

# **Altruistische Bestrafung**

Inauguraldissertation  
zur  
Erlangung des Doktorgrades  
der  
Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der  
Universität zu Köln

2007

vorgelegt

von

Diplom Volkswirt Franz Nikolaus Gresser

aus

Binningen (CH)

Referent: Prof. Dr. Detlef Fetchenhauer

Korreferent: Prof. Dr. Günter Wiswede

Tag der Promotion: 13.07.2007

## **Dank**

Zunächst bedanke ich mich bei Herrn Prof. Dr. Detlef Fetchenhauer, dem Betreuer meines Dissertationsvorhabens. Er hatte stets ein offenes Ohr und stand mir mit guten und motivierenden Ratschlägen zur Seite. Herrn Prof. Dr. Günter Wiswede und Herrn Prof. Dr. Lorenz Fischer danke ich dafür, dass sie mein Interesse an der Wirtschafts- und Sozialpsychologie geweckt haben und mir die Möglichkeit eingeräumt haben, in diesem spannenden Bereich tätig zu sein.

Mein Dank geht auch an Thomas Schlösser und Frau Prof. Dr. Gabriele Jacobs-Belschak für viele fruchtbare Diskussionen sowie an meine Kollegen Alexandra Haferkamp und Dr. Thomas Marchlewski, die mir bei großen und kleinen Nöten halfen.

Zuletzt und ganz besonders herzlich bedanke ich mich bei meinen Eltern und meiner Freundin. Meine Eltern haben mich stets unterstützt und durch ihre konstruktiven kritischen Kommentare diese Dissertation verbessert. Meine Freundin Caroline Morich gibt mir den nötigen emotionalen Rückhalt.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>IV</b>
<b>Tabellenverzeichnis .....</b>	<b>V</b>
<b>Formelverzeichnis .....</b>	<b>V</b>
<b>1. Einleitung .....</b>	<b>1</b>
<b>2. Altruistische Bestrafung.....</b>	<b>6</b>
<b>2.1. Biologische vs. psychologische Definition von Altruismus.....</b>	<b>6</b>
<b>2.2. Altruistische Bestrafung: Ist Bestrafung überhaupt ein altruistischer Akt? .....</b>	<b>7</b>
2.2.1. <i>Altruistische Bestrafung als Rätsel der Evolution.....</i>	<i>7</i>
2.2.2. <i>Kosten der altruistischen Bestrafung.....</i>	<i>8</i>
2.2.3. <i>Beispiele für Reale Situationen, in denen altruistische Bestrafung vorkommt.....</i>	<i>8</i>
<b>3. Erklärungen für altruistische Bestrafung.....</b>	<b>10</b>
<b>3.1. Proximate Erklärungen für Altruistische Bestrafung ...</b>	<b>10</b>
3.1.1. <i>Gerechtigkeitspsychologische Erklärungen.....</i>	<i>11</i>
3.1.2. <i>Soziale Normen.....</i>	<i>16</i>
3.1.3. <i>Persönlichkeit und Altruistische Bestrafung: Die Ungerechtigkeitssensibilität .....</i>	<i>18</i>
3.1.4. <i>Emotionen.....</i>	<i>24</i>
3.1.5. <i>Ökonomische Modelle: Soziale Präferenzen .....</i>	<i>29</i>
3.1.6. <i>Zusammenfassung der proximatzen Erklärungen .....</i>	<i>32</i>
<b>3.2. Ultimate Erklärungen für Altruistische Bestrafung.....</b>	<b>33</b>
3.2.1. <i>Verwandtschaftsaltruismus: Blut ist dicker als Wasser..</i>	<i>33</i>
3.2.2. <i>Reziproker Altruismus: Wie du mir, so ich dir .....</i>	<i>34</i>
3.2.3. <i>Indirekte Reziprozität: Wie du ihm, so ich dir.....</i>	<i>36</i>

3.2.4.	<i>Strong Reciprocity, Gruppenselektion, Mehrebenenselektion und Ko-Evolution von Genen und Kultur .....</i>	37
3.2.5.	<i>Das Commitment-Modell .....</i>	39
3.2.6.	<i>Alte Verhaltensweisen (fälschlich) angewandt auf neue Situationen .....</i>	41
3.2.7.	<i>Zusammenfassung der ultimativen Erklärungen.....</i>	42
<b>4.</b>	<b>Altruistische Bestrafung in experimentellen Spielen .....</b>	<b>44</b>
<b>4.1.</b>	<b>Spieltheorie und Spiele.....</b>	<b>44</b>
4.1.1.	<i>Spieltheorie als Sprache und Instrument experimenteller Ökonomen .....</i>	44
4.1.2.	<i>Grundlegende Charakteristika experimenteller Spiele.....</i>	46
<b>4.2.</b>	<b>Second-Party-Punishment (SPP).....</b>	<b>51</b>
4.2.1.	<i>Ultimatum-Spiel (Ultimatum Game) .....</i>	52
4.2.2.	<i>Diktator-Spiel (Dictator Game).....</i>	59
4.2.3.	<i>Öffentliches-Gut-Spiel mit Bestrafung (public goods game).....</i>	62
4.2.4.	<i>Gefangenen-Dilemma-Spiel (prisoners dilemma).....</i>	68
4.2.5.	<i>Vertrauens-Spiel (Trust Game) bzw. Investitions-Spiel ....</i>	73
4.2.6.	<i>Moonlighting-Spiel .....</i>	76
4.2.7.	<i>Drei-Personen-Ultimatum-Spiel („Mixed Game“).....</i>	79
4.2.8.	<i>Power-To-Take-Spiel.....</i>	82
<b>4.3.</b>	<b>Third-Party-Punishment: Bestrafung durch einen unbeteiligten Dritten .....</b>	<b>84</b>
4.3.1.	<i>Drei-Personen-Bestrafungs-Spiel.....</i>	85
4.3.2.	<i>Gefangenen-Dilemma mit Bestrafung durch Dritte .....</i>	87
<b>4.4.</b>	<b>Vergleich von Second- und Third-Party-Punishment ...</b>	<b>90</b>
<b>4.5.</b>	<b>Einflüsse auf Art und Höhe der Bestrafung.....</b>	<b>95</b>
4.5.1.	<i>Kosten der Bestrafung .....</i>	95
4.5.2.	<i>Höhe der Beträge .....</i>	96
4.5.3.	<i>Strategiemethode.....</i>	96
4.5.4.	<i>Anonymität.....</i>	98
4.5.5.	<i>Intentionen (Spielpartner Mensch vs. Computer).....</i>	99

4.5.6.	<i>Gruppenzugehörigkeit und Gruppengröße .....</i>	<i>100</i>
4.5.7.	<i>Kommunikation .....</i>	<i>101</i>
<b>4.6.</b>	<b>Bösartige Bestrafung .....</b>	<b>102</b>
<b>4.7.</b>	<b>Kritik an experimentellen Spielen zur Messung altruistischer Bestrafung .....</b>	<b>102</b>
4.7.1.	<i>Situation künstlich, kontextfreie Darstellung .....</i>	<i>103</i>
4.7.2.	<i>Geldbeträge meist zu gering.....</i>	<i>104</i>
4.7.3.	<i>Spiele werden nicht verstanden.....</i>	<i>104</i>
<b>5.</b>	<b>Experimentelle Untersuchung altruistischer Be- strafung .....</b>	<b>107</b>
<b>5.1.</b>	<b>Ungerechtigkeitssensibilität, negative Emotionen und altruistische Bestrafung.....</b>	<b>107</b>
5.1.1.	<i>Studie 1 .....</i>	<i>107</i>
5.1.2.	<i>Studie 2 .....</i>	<i>119</i>
5.1.3.	<i>Gesamtdiskussion der Studien 1 und 2.....</i>	<i>134</i>
<b>5.2.</b>	<b>Die Wirksamkeit von altruistischer Bestrafung: Berücksichtigen Individuen die Bestrafungsneigung ihrer Mitmenschen? .....</b>	<b>138</b>
5.2.1.	<i>Studie 3 .....</i>	<i>140</i>
5.2.2.	<i>Studie 4 .....</i>	<i>151</i>
5.2.3.	<i>Gesamtdiskussion der Studien 3 und 4.....</i>	<i>158</i>
<b>6.</b>	<b>Schlussfolgerungen und Ausblick .....</b>	<b>163</b>
<b>7.</b>	<b>Literatur .....</b>	<b>169</b>
<b>8.</b>	<b>Anhang .....</b>	<b>184</b>

## Abbildungsverzeichnis

Abbildung 1: Kooperation im Öffentliches-Gut-Spiel ohne und mit Bestrafung (Fehr & Gächter, 2000, S. 989) .....	68
Abbildung 2: Aufbau der Bestrafung beim Vergleich zwischen Second- und Third-Party-Punishment im Diktator-Spiel (Fehr & Fischbacher, 2004b, S. 75) .....	92
Abbildung 3: Ausgaben für die Bestrafung in Abhängigkeit vom Anteil für den Empfänger (Fehr & Fischbacher, 2004b, S. 78, eigene Übersetzung der Achsenbeschriftung) .....	93
Abbildung 4: Erwarteter Gewinn des Diktators in Abhängigkeit vom Anteil für den Empfänger (Fehr & Fischbacher, 2004b, S. 79, eigene Übersetzung der Achsenbeschriftung) .....	93
Abbildung 5: Aufbau der Bestrafung beim Vergleich zwischen Second- und Third-Party-Punishment im Gefangenen-Dilemma (Fehr & Fischbacher, 2004b, S. 81) .....	94
Abbildung 6: Angenommener Wirkungszusammenhang US - negative Emotionen – Entscheidung mit standardisierten ( $\beta$ ) und unstandardisierten (B) Regressionsgewichten und Standardfehlern (SE) (*: $p < .05$ ; **: $p < .01$ ) .....	115
Abbildung 7: Aufteilungsentscheidung der simulierten Person A (Studie 2) .....	123
Abbildung 8: Angenommener Wirkungszusammenhang US - negative Emotionen – Entscheidung mit standardisierten ( $\beta$ ) und unstandardisierten (B) Regressionsgewichten und Standardfehlern (SE) (*: $p < .05$ ; **: $p < .01$ ) .....	127
Abbildung 9: Erwartungswert der Aufteilung für den Aufteilenden in Abhängigkeit vom Angebot an den Entscheider und den Dummy im Drei-Personen-Ultimatum-Spiel .....	149
Abbildung 10: Erwartungswert der Auszahlung für den Aufteilenden in Abhängigkeit vom Angebot an den Empfänger im Drei-Personen-Bestrafungs-Spiel.....	157

## **Tabellenverzeichnis**

Tabelle 1: Strafen im Gefangenen-Dilemma (S1: Spieler 1; S2: Spieler 2) .....	69
Tabelle 2: Wahrscheinlichkeit mit der ein Angebot akzeptiert wurde .....	148

## **Formelverzeichnis**

Formel 1: Berechnung der Signifikanz des Mediationseffektes (a, b: unstandardisierte Regressionskoeffizienten; $s_a$ , $s_b$ : Standardfehler, siehe Abbildung weiter oben) .....	116
---	-----



„Werden die Bösen bestraft, so ist dies ohne Frage gerecht. Gehen sie aber straflos aus, so widerspricht dies der Gerechtigkeit. Nicht wahr?“ – „Unleugbar!“ – „Auch das aber ist unleugbar, dass alles Gerechte, also auch die Bestrafung der Bösen, auch gut, alles Ungerechte, also ihre Straflosigkeit, aber etwas Böses ist!“

(Boethius: Die Tröstungen der Philosophie)

## 1. Einleitung

Gesetze regeln weite Teile der menschlichen Interaktion. Vieles, was Gesetze nicht erfassen, wird durch oft informelle soziale Normen, z.B. der Fairness und der Gerechtigkeit, bestimmt. Die meisten Menschen vertrauen auf diese Normen und halten sich daran. Häufig kommt es aber zu Dilemma-Situationen, in denen ein Individuum vor einer schwierigen Entscheidung steht: eine Normverletzung ist individuell lukrativ, für die Gemeinschaft ist sie jedoch schädlich. Wie können Normen in diesem Spannungsfeld von Eigennutz und Gemeinwohl dennoch durchgesetzt werden? Setzt man eigennützig handelnde Individuen voraus, wie es die Ökonomie grundsätzlich tut, kommt man zu dem Schluss, dass es in dieser Art von Dilemmasituationen üblicherweise nur ein stabiles Gleichgewicht in der Gesellschaft geben kann: Niemand hält sich an die Normen, weil das, unabhängig vom Verhalten aller anderen, die beste Strategie für das Individuum ist. Ein solches Treiben hätte verheerende Folgen für die Gesellschaft. Aber lässt sich ein allgegenwärtiger Normbruch in der Gesellschaft beobachten? Nein, gelegentlich wird zwar über den Verfall gesellschaftlicher Normen geklagt und dennoch gibt es in den meisten menschlichen Gesellschaften – insbesondere in den ökonomisch erfolgreichen (Fetkenhauer & van der Vegt, 2001) – ein hohes Maß an Vertrauen, Normeinhaltung, Altruismus und Kooperation, das weit über dem liegt, was man aus den Voraussagen der Ökonomie und der rationalen Entscheidungstheorie ableiten würde.

Wie wird aber dieses hohe Maß an Normeinhaltung aufrechterhalten, wenn sich zeigen lässt, dass es für ein einzelnes Individuum vorteilhaft sein kann, sich nicht an die Normen zu halten? Neben anderen Mechanismen, die der Beachtung von Normen dienen, kann die Bestrafung von Normbrechern durch gewöhnliche Mitglieder der Gesellschaft (im Gegensatz zu hoheitlichen Autoritäten) dieses soziale Problem lösen. Unter Bestrafung ist hier jede Form von negativer Sanktion zu verstehen, die absichtlich gegen eine Person verhängt wird, von der wahrgenommen wird, dass sie gegen ein Gesetz, eine Regel, eine Norm oder eine Erwartung verstoßen hat (Miller & Vidmar, 1981).

Bestrafung nach der Verletzung von Verteilungs- und Kooperationsnormen lässt sich im Alltag beobachten und wurde auch experimentell vielfach nachgewiesen (vgl. z.B. Güth, Schmittberger & Schwarze, 1982; Kahneman, Knetsch & Thaler, 1986a; Fehr & Gächter, 2002, 2000). Da die Bestrafung eines Normbrechers in der Regel aber mit Kosten für den Bestrafenden verbunden ist – im besten Fall sind das monetäre Kosten oder Zeit, im schlimmsten Fall die Gesundheit oder das Leben – stellt sich die Frage: Warum nehmen Menschen diese Kosten auf sich, um andere zu bestrafen?

Verschiedene Fachrichtungen befassen sich unter unterschiedlichen Namen mit dieser Form der Bestrafung: In der Ökonomie wird es als „altruistische Bestrafung“ (vgl. z.B. Fehr & Gächter, 2002), in der soziologischen Forschung als „norm enforcement“ (vgl. z.B. Horne & Cutlip, 2002) und in der Evolutionsbiologie als „strong reciprocity“ (vgl. z.B. Gintis, 2000) bezeichnet.<sup>1</sup> In der Psychologie setzt man sich mit diesem Phänomen unter dem Stichwort „retributive Gerechtigkeit“ auseinander (vgl. z.B. Hogan & Emler, 1981; Miller & Vidmar, 1981; Vidmar, 2001) und auch mit der Forschung im Bereich der Rache und Vergeltung gibt es Überschneidungen (vgl. Gollwitzer, 2004, zum Zusammenhang von Rache, Vergeltung und Strafe).

---

<sup>1</sup> Die Konzepte „strong reciprocity“ und „norm enforcement“ umfassen im Gegensatz zur altruistischen Bestrafung auch Belohnungen für kooperatives Verhalten und Normkonformität.

Zur Untersuchung dieses Phänomens nutzen die verschiedenen Disziplinen experimentelle Spiele. Das Instrument der experimentellen Spiele stellt eine gemeinsame Sprache zur Verfügung, die die Kommunikation zwischen verschiedenen Fachrichtungen ermöglicht. Leider findet trotzdem wenig Austausch zwischen den Disziplinen statt. Diese Arbeit stellt – über die konkreten Untersuchungen hinaus – den Versuch dar, verschiedene Ideen aus unterschiedlichen Fachrichtungen miteinander zu kombinieren und soll so auch eine gemeinsame Diskussionsgrundlage bieten.

In dieser Arbeit wird der Terminus „altruistische Bestrafung“ verwendet, da er sich in der momentanen, stark von der experimentellen Ökonomie geprägten wissenschaftlichen Diskussion durchgesetzt hat und es hier ausschließlich um Bestrafung geht. Außerdem wird durch diesen Begriff der Aspekt der Kosten für den Bestrafenden sowie der durch die Bestrafung entstehende Nutzen für die Gesellschaft hervorgehoben.

Ist die bestrafende Person selbst von der Untat betroffen, ist leicht einzusehen, dass sie durch Bestrafung und Vergeltung versuchen wird, Gerechtigkeit wieder herzustellen (Second-Party-Punishment). Schwieriger zu begründen ist, dass unbeteiligte Dritte strafend in ein Geschehen eingreifen, von dem sie nicht direkt betroffen sind (Third-Party-Punishment). Jedoch lässt sich auch diese Form der Bestrafung durch unbeteiligte Dritte in der Realität und in Experimenten feststellen (Fehr & Fischbacher, 2004b). Die positiven Folgen dieses Verhaltens für die Gesellschaft sind weit reichend, denn wenn sich aus der Bestrafungsneigung Dritter eine hinreichend große Bestrafungsdrohung für Normverletzungen ergibt, vervielfacht eine Gesellschaft ihre Möglichkeiten, gesellschaftliche Normen durchzusetzen (Camerer & Fehr, 2004; Fehr, Fischbacher & Gächter, 2002). Für die Ökonomie war dieses Verhalten lange Zeit ein Rätsel, da es mit dem Bild des Homo Oeconomicus – dem streng rational und eigennützig handelnden Individuum – nicht vereinbar ist.

Dieses Unverständnis rührt zum Teil aus den zu engen Annahmen der ökonomischen Modelle her, die meist Eigennutz als universelles und einziges Motiv annehmen. Verschiedentlich wurde gezeigt, dass der Egoismus der Menschen überschätzt wird (z.B. Miller & Ratner, 1998). Auch Gerechtigkeitsmotive spielen neben Eigennutzmotiven eine zentrale Rolle (Montada, 1998). Darüber hinaus spielen in der ökonomischen Theorie die Persönlichkeit der Handelnden und deren Emotionen keine Rolle, in der Realität aber sehr wohl. Für verschiedene Persönlichkeitsdispositionen konnte ein Zusammenhang mit dem Bestrafungsverhalten nachgewiesen werden (Fetchenhauer & Huang, 2004; Brandstätter & Güth, 2002; Brandstätter & Königstein, 2001).

Auch bei Berücksichtigung des Gerechtigkeitsmotivs und der Persönlichkeit bleibt jedoch die Tatsache bestehen, dass kooperative und strafende Individuen in jeder einzelnen Interaktion einen individuellen (materiellen) Nachteil gegenüber solchen Mitmenschen haben, die sich nicht an Normen halten und die nicht bestrafen. Eine rein materielle Kosten-Nutzen-Analyse würde also gegen die altruistische Bestrafung eines Normbrechers sprechen. Dieser Mangel kann durch die Einbeziehung moralischer Emotionen behoben werden. Negative Emotionen wie Ärger, treten als Reaktion auf die Verletzung einer Norm auf. Solche moralischen Emotionen können die Kosten und Nutzen der Bestrafung dramatisch verändern und Personen, die moralische Emotionen äußern, haben unter Umständen sogar einen evolutionären (Fitness-)Vorteil gegenüber Menschen, die dies nicht tun, da sie häufiger als Interaktionspartner ausgesucht werden (Frank, 1988). Während die psychologische Forschung schon seit einiger Zeit den Zusammenhang zwischen negativen Emotionen und der Bestrafung kennt (für einen Überblick siehe Vidmar, 2001), wurden solche Zusammenhänge in der Ökonomie nach Adam Smiths „The Theory of Moral Sentiments“ aus dem Jahr 1759 weitgehend „vergessen“ und erst in jüngerer Zeit „wiederentdeckt“ (siehe z.B. Fehr & Gächter, 2002).

Mit den Fragen: „Warum bestrafen Menschen altruistisch?“ und „Rechnen Menschen damit, dass ihre Mitmenschen sie bestrafen könnten?“ beschäftigt sich der empirische Teil der vorliegenden Arbeit. Wenn Menschen die Bestrafung durch andere nicht antizipieren würden, könnte die Bestrafung keine verhaltenssteuernde Wirkung entfalten und würde lediglich zur Vernichtung von Ressourcen führen, denn den Kosten der Bestrafung stände dann kein (zukünftiger) gesellschaftlicher Nutzen gegenüber. Zur Beantwortung der Fragen werden in der vorliegenden Arbeit verschiedene Erklärungen – hauptsächlich aus der Psychologie und der experimentellen Ökonomie – einander gegenübergestellt und diskutiert. Darüber hinaus werden mögliche Antworten auf ihre Vereinbarkeit mit evolutionären Theorien geprüft.

Im folgenden Abschnitt wird der Begriff der altruistischen Bestrafung genauer spezifiziert. Anschließend werden verschiedene Erklärungen für altruistische Bestrafung, die sich in ihrer Nähe zum tatsächlichen Verhalten unterscheiden, vorgestellt und diskutiert. Danach wird in einem allgemeinen Überblick auf die Messung der altruistischen Bestrafung in experimentellen Spielen eingegangen. Im Anschluss daran werden die vier eigenen empirischen Untersuchungen dargestellt, die genauer auf Ursachen und Wirkungen der altruistischen Bestrafung eingehen. Die Studien 1 und 2 befassen sich mit Ursachen altruistischen Bestrafens, nämlich mit der Persönlichkeit der Bestrafenden und den Emotionen, die Bestrafung motivieren und auslösen. Die Studien 3 und 4 untersuchen die Frage, ob Menschen die Bereitschaft ihrer Mitmenschen zur altruistischen Bestrafung antizipieren und ihre Handlungen entsprechend anpassen.

## **2. Altruistische Bestrafung**

„Homo homini lupus est“: Ist das menschliche Leben ein ständiger Kampf oder verhalten wir uns (auch) freundlich und hilfsbereit gegenüber unseren Mitmenschen? Zunächst muss geklärt werden, was unter „Altruismus“ zu verstehen ist, und es werden die biologische und die psychologische Sichtweise auf altruistisches Verhalten erläutert.

### **2.1. Biologische vs. psychologische Definition von Altruismus**

Evolutionäre Biologen verstehen unter „Altruismus“ ein Verhalten eines Organismus A, das die reproduktive Fitness eines Organismus B erhöht, während es die reproduktive Fitness von A verringert (Sigmund & Hauert, 2002). Ähnlich, jedoch in einer anderen Sprache definieren Camerer und Fehr (2004) „Altruismus“ als ein Verhalten, bei dem der Handelnde 1. ohne Ansehen vorheriger eigener Handlungen oder der Handlungen Anderer 2. Kosten auf sich nimmt und 3. den Ertrag einer anderen Person steigert. Altruismus hat in diesem Sinne keine historische und keine strategisch-zukünftige Komponente. Die biologische Sichtweise beschränkt sich auf die Kosten und Nutzen, die den verschiedenen Akteuren entstehen. Motive und Emotionen, welche die Handlungen begleiten, werden nicht berücksichtigt.

Im Gegensatz hierzu bezieht der psychologische Altruismus auch die Motive und die während der Handlung entstehenden Gefühle mit ein. Eine Handlung, die Kosten beim Handelnden hervorruft und zu positiven Erträgen bei einer anderen Person führt, wird im psychologischen Sinn nur dann als altruistisch bezeichnet, wenn der Handelnde aus nicht-eigennützigen Motiven handelt (Bierhoff, 2002). Dies sicher zu stellen, dürfte nahezu unmöglich und zudem empirisch schwer prüfbar sein. Wenn ein Mensch z.B. 10 Euro spendet, um

sein Ansehen bei anderen zu erhöhen, ist das im biologischen Sinne zwar altruistisch, im psychologischen Sinne jedoch nicht. Während das Motiv schwer zu ermitteln ist, kann die Spende eindeutig gemessen werden. Im weiteren Verlauf dieser Arbeit wird aus Gründen der besseren Operationalisierbarkeit der biologischen Definition von Altruismus gefolgt.

## **2.2. Altruistische Bestrafung: Ist Bestrafung überhaupt ein altruistischer Akt?**

Für den Fall einer Belohnung ist unmittelbar einleuchtend, dass es sich um einen altruistischen Akt handelt. Die Belohnung verringert die Ressourcen des Belohnenden und erhöht die Ressourcen des Belohnten. Dass es sich bei altruistischer Bestrafung um einen altruistischen Akt handelt, lässt sich aus der (biologischen) Definition von Altruismus ableiten: Wer unkooperative Andere unter Kosten bestraft, hilft denen, die von einer zukünftigen Einhaltung der Kooperationsnorm durch den Bestraften profitieren (O'Gorman, Wilson & Miller, 2005). Eine bestrafende Person nimmt also Kosten auf sich und nützt anderen Menschen durch ihr Verhalten.

### **2.2.1. Altruistische Bestrafung als Rätsel der Evolution**

Per definitionem ist altruistische Bestrafung (und Altruismus allgemein) eine Verhaltensweise, welche die eigene Fitness verringert und die eines anderen Wesens erhöht. Hieraus leitet sich direkt die Frage ab: Warum wurden solche Verhaltensweisen nicht im Lauf der menschlichen Evolution gelöscht (Fetchenhauer & Bierhoff, 2004)? In der Tat fällt es innerhalb vieler Theorien der Evolution schwer, altruistisches Verhalten zu erklären (für einen Überblick siehe Fetchenhauer & Bierhoff, 2004). Im Rahmen der ultimativen Erklärungen für altruistische Bestrafung soll hierauf weiter unten genauer eingegangen werden.

### **2.2.2. Kosten der altruistischen Bestrafung**

Es sind verschiedene Arten von Kosten denkbar. Bestrafende müssen in jedem Fall Zeit aufwenden, evtl. sind sie gezwungen, in eigentlich von ihnen nicht gewünschter Art und Weise öffentlich Stellung zu beziehen. Im Extremfall kann es sein, dass sie hohe Risiken eingehen und evtl. sogar ihre Gesundheit oder ihr Leben verlieren. Eine ganz offensichtliche Form von Kosten ist der Nutzen- oder Geldverlust, den der Bestrafende durch die Zurückweisung eines unfairen Angebots eines Anderen erleidet. Anstatt einen kleinen Anteil bei einem unfairen Handel zu erhalten, bekommt er nichts. Fehr und Fischbacher (2004a) nennen unter anderem auch das Risiko der Vergeltung, emotionale Anspannung und den möglichen Verlust einer Beziehung als Kosten der Bestrafung.

Bemerkenswert ist, dass in einer Gesellschaft die Kosten der Bestrafung mit steigender Normeinhaltung sinken: Wenn die zu bestrafende Normverletzung nicht mehr auftritt, weil alle sich normkonform verhalten, muss nicht bestraft werden und die Kosten gehen auf Null zurück (Camerer & Fehr, 2004). Hierin ist auch ein Kostenvorteil gegenüber der Belohnung zu sehen. Aufwendungen für die Belohnung eingehaltener Normen steigen mit dem Grad der Normkonformität. Welchen Einfluss die Höhe der Kosten auf das Bestrafungsverhalten haben kann, wird weiter unten näher geklärt.

### **2.2.3. Beispiele für Reale Situationen, in denen altruistische Bestrafung vorkommt**

Wenn Menschen in Gruppen zusammen arbeiten, in denen es bestimmte Normen und Ansichten über den fairen Beitrag jedes einzelnen Mitglieds gibt, wird ein Mitglied, das von diesen Normen (negativ) abweicht, von den anderen bestraft (Vidmar, 2001). Diese Strafe kann z.B. in der Ausgrenzung aus der Gruppe bestehen (vgl. z.B. Masclet, 2003). Das ist für beide Seiten mit Kosten verbunden, denn aus der Sicht der Gruppe geht ihr der geringe Anteil, den das unfaire Mitglied bisher leistete, verloren und er muss durch die restlichen Mitglieder



der Gruppe zusätzlich erbracht werden. Für das unfaire Mitglied sind die Kosten allerdings ungleich höher, denn ihm gehen die anteiligen Beiträge aller anderen Gruppenmitglieder verloren, wenn es ausgeschlossen wird.

Skarlicki und Kulik (2005) berichten davon, dass es in Firmen, die Mitarbeiter entlassen haben, zu Bestrafungen der Firma durch die im Unternehmen verbliebenen Mitarbeiter kommen kann. Dieses Beispiel für die Bestrafung durch eine dritte Partei verdeutlicht die großen Auswirkungen der Bereitschaft zur altruistischen Bestrafung.

Ein weiteres Beispiel für altruistische Bestrafung stellt der Wechsel des Geschäfts aufgrund von Fairnessüberlegungen dar. Kahneman, Knetsch und Thaler (1986b) berichten, dass 69% der Befragten in einer Telefonumfrage zu einem weiter entfernten Geschäft wechseln würden, wenn sie erführen, dass der bequemer zu erreichende Laden seine älteren Mitarbeiter diskriminiert.

Darunter fällt auch der Nicht-Kauf favorisierter Turnschuhe einer bestimmten Firma, weil diese unter unfairen Bedingungen in der Dritten Welt fertigen lässt. Der Kunde muss auf den (Mehr-)Nutzen, den genau diese Turnschuhe ihm stiften, verzichten (Kosten für den Bestrafenden). Die Firma allerdings verliert einen Kunden, und zusätzlich kann es sein, dass der Kunde sogar noch andere überzeugen kann, es ihm gleich zu tun.

### **3. Erklärungen für altruistische Bestrafung**

#### **3.1. Proximate Erklärungen für Altruistische Bestrafung**

Proximate Erklärungen sind Erklärungen, die direkt oder nahe am untersuchten Gegenstand ansetzen. Sie sind Antworten auf die Frage: Wie funktioniert ein bestimmtes System? Proximate Modelle versuchen entsprechend, das beobachtbare Verhalten einer Person durch ihre Motive (oder Emotionen) unmittelbar zu erklären. Von ihnen abzugrenzen sind ultimate Erklärungen, die auf die evolutionäre Entwicklung der Menschheit zurückgreifen, um die Entstehung von Motiven oder Emotionen als Verhaltensauslöser zu begründen. Sie sind Antworten auf die Frage: Warum hat sich dieses System so entwickelt? (Die Unterscheidung geht auf Tinbergen, 1963 zurück; für eine Erläuterung im Rahmen der evolutionären Psychologie siehe Gaulin & McBurney, 2004). Auf die evolutionäre Entwicklung und die ultimatsten Erklärungen für altruistische Bestrafung wird im nächsten Kapitel eingegangen. Zunächst soll das Verhalten in den experimentellen Spielen unmittelbar erklärt werden.

Als erstes werden psychologische Prinzipien distributiver Gerechtigkeit erläutert. Im Anschluss daran wird auf die Rolle (sozialer) Normen eingegangen. Daran anschließend wird beschrieben, welche Persönlichkeitsmerkmale einen Einfluss auf das Bestrafungsverhalten haben, wobei ein Schwerpunkt auf die Ungerechtigkeitssensibilität gelegt wird. Da verschiedene Erklärungen Emotionen als wichtige Bestandteile beinhalten, wird dann auf die Rolle der Emotionen bei Bestrafungen eingegangen. Danach werden ökonomische Modelle vorgestellt, die das Verhalten in sozialen Dilemmasituationen durch verschiedene soziale Präferenzen erklären.

### 3.1.1. Gerechtigkeitspsychologische Erklärungen

Die Gerechtigkeitspsychologie befasst sich damit, was Menschen als gerecht bzw. ungerecht empfinden und wie sie darauf reagieren. Hierbei werden sowohl Emotionen als auch Kognitionen und das Verhalten berücksichtigt.

Verschiedene Aspekte der Gerechtigkeit können voneinander abgegrenzt werden (Gollwitzer, Schmitt, Fetschenhauer, Baumert & Schlösser, in Druck):

1. Distributive Gerechtigkeit (Verteilungsgerechtigkeit): Z.B. Wie soll ein bestimmter Geldbetrag zwischen zwei Personen aufgeteilt werden?
2. Prozedurale Gerechtigkeit (Verfahrensgerechtigkeit): Z.B. Nach welchen Regeln soll eine Beförderung eines Mitarbeiters vorgenommen werden?
3. Retributive Gerechtigkeit (Vergeltungsgerechtigkeit): Z.B. (Wie) soll eine Person für eine ungerechte Handlung bestraft werden?

Für die vorliegende Arbeit sind insbesondere distributive Gerechtigkeit (Wie sollen Güter/Lasten gerecht verteilt werden?) und retributive Gerechtigkeit (Mit welcher Sanktion sollte auf einen Normbruch reagiert werden?) relevant (Gollwitzer et al., in Druck; Hogan & Emler, 1981) und sie werden daher nachfolgend genauer erläutert.

*Distributive Gerechtigkeit: Wie soll aufgeteilt werden?*

Bei der Frage, wie man Güter und Lasten gerecht verteilen soll, können verschiedene Gerechtigkeitsprinzipien walten, über deren genaue Zahl Uneinigkeit herrscht. Einhelligkeit herrscht aber darüber, dass mindestens drei Prinzipien voneinander unterschieden werden müssen (Schmitt, 1996, S. 3, spricht von den "big three"):

1. Das Leistungsprinzip (equity): Das Verhältnis von Einsatz und Ertrag ist für alle gleich.

2. Das Gleichheitsprinzip (parity): Alle erhalten das Gleiche, unabhängig von ihrem Einsatz oder ihrer Person.
3. Das Bedürfnisprinzip (need): Jeder erhält entsprechend seiner Bedürftigkeit.

Herrscht Einigkeit darüber, welches Prinzip angewendet werden sollte und werden Erträge und Lasten entsprechend aufgeteilt, wird das Resultat als gerecht wahrgenommen.

In ökonomischen Beziehungen wird meist Equity als angemessenes Verteilungsprinzip betrachtet (s. u. bei Inequity-Aversion). Innerhalb einer Familie oder in Freundes-/ Liebesbeziehungen wird eher das Gleichheits- bzw. Bedürfnisprinzip vorherrschen. Welches Prinzip als das angemessene angesehen wird, hängt maßgeblich von der Interpretation der Beziehung bzw. der Situation ab (Lerner, 1977).

Die Equity-Theorie (Walster, Walster & Berscheid, 1978; Adams, 1965; Homans, 1961) hat früh zur Etablierung der Gerechtigkeitspsychologie als eigenes Feld beigetragen und beeinflusst mittlerweile auch ökonomische Theorien. Sie behandelt die Verteilungsgerechtigkeit. Im Vordergrund steht die Leistungsgerechtigkeit, gemessen über das Verhältnis von Leistungen (Input) und Erträgen (Output). Ist dieses Verhältnis für verschiedene Personen gleich, herrscht „equity“. Eine wahrgenommene Ungleichheit des Leistungs-Ertrags-Verhältnisses löst gemäß der Theorie einen unangenehmen Spannungszustand aus, der auf Beseitigung der Ungleichheit drängt, wobei eine Besserstellung zu geringerem Unbehagen führt als eine Schlechterstellung. Die Equity-Theorie geht außerdem davon aus, dass Individuen danach trachten, ihre Erträge zu maximieren, unterstellt also wie die ökonomischen Theorien das Eigennutzmotiv als grundlegend.

*Retributive Gerechtigkeit: Wer sollte wie bestraft werden?*

Retributive Gerechtigkeit befasst sich mit der Bestrafung für erlittenes Unrecht. Im Mittelpunkt steht die Wiederherstellung der Ge-

rechtheit (Hogan & Emler, 1981; Miller & Vidmar, 1981; Vidmar, 2001; Gollwitzer et al., in Druck).

Bei der Frage der Bestrafung muss die Frage des Motivs bzw. des Ziels gestellt werden. Vidmar (2001) unterscheidet zwischen zwei Zielen:

1. Beeinflussung des Verhaltens: Die Beeinflussung des Verhaltens ist einem utilitaristischen Verständnis von Gerechtigkeit zuzuordnen. Bei der Verhängung einer Strafe, die das Verhalten des Normbrechers beeinflussen soll, geht es nicht um den moralischen Charakter des Normbrechers und nicht darum, einem moralischen Gesetz genüge zu tun. Es soll allein sichergestellt werden, dass sich der Normbrecher jetzt und in Zukunft an die Norm hält.
2. Vergeltung (retribution): Bei Vergeltung geht es um die moralischen Aspekte des Normbruchs. Die Bestrafung wird motiviert durch den Normbruch. Sie leitet sich aus der Norm (aus der höheren moralischen Regel) ab und nicht aus der Verhinderung zukünftigen Fehlverhaltens.

Der erste Ansatz entspricht einem utilitaristischen Gerechtigkeitsverständnis, das seine Rechtfertigung für Strafen aus zukünftigen Konsequenzen der Bestrafung ableitet.

Der Zweite Ansatz entspricht der deontischen (griechisch deon = Pflicht) Auffassung von Gerechtigkeit (siehe Cropanzano, Goldman & Folger, 2003, sowie; Turillo, Folger, Lavelle, Umphress & Gee, 2002, für deontic justice). Hier wird dem Nutzengedanken ein eigenständiges Gerechtigkeitsmotiv gegenübergestellt. Gerechtigkeit wird hier nicht als Mittel zum Zweck gesehen, sondern selbst als Zweck (Turillo et al., 2002; Montada, 1998; Lerner, 1977). Die deontische Sichtweise besagt, dass Menschen aufgrund einer moralischen Verpflichtung das Richtige tun. Grundlage für moralisches Verhalten sind übergeordnete Regeln oder Prinzipien. Diese Ansicht hat Kant (1785, S. 2. Abschnitt, 418 f. Hervorhebung im Original) in seinem „Kategori-

schen Imperativ“ formuliert: „Der kategorische Imperativ ist also nur ein einziger, und zwar dieser: *handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.*“

Dieser Punkt wird auch deutlich, wenn Westermarck über die moralischen Emotionen folgendes sagt: “A moral judgement always has the character of disinterestedness. When pronouncing an act of good or bad, I mean that it is so quite independently of any reference it might have to me personally.” (Westermarck, 1932, S. 90, zitiert nach Heider, 1958, S.272). Menschen spüren/wissen demnach, was in einer bestimmten Situation das moralisch Richtige ist und handeln entsprechend, auch wenn dadurch eigennützige Ziele evtl. gefährdet werden (Turillo et al., 2002).

Auch Fetschenhauer (2001) kritisiert, dass insbesondere in der Ökonomie Moral häufig lediglich als Epiphänomen betrachtet wird. Auch er vertritt im Prinzip eine deontische Sichtweise, wenn er ein Modell vorstellt, in dem Moral als Filter vor Kosten-Nutzen-Analysen fungiert. Wiswede (2001) verweist allerdings darauf, dass ein solches Modell eher bei sehr starken Moralüberzeugungen gelten dürfte. Auch Kohlberg (1968/1995, S. 20) nimmt eine vermittelnde Position ein: „Moralische Urteile und Entscheidungen sind in allen Kulturen eine Mischung von Urteilen, die unter dem Gesichtspunkt der Konsequenzen für den individuellen menschlichen Nutzen oder aber unter dem der konkreten kategorischen Sozialregeln gefällt werden“. Meiner Ansicht nach wird es immer von der Persönlichkeit und der jeweiligen perzipierten Situation abhängen, wie sich eine Person verhält. Ein rein regelgeleitetes moralisches Verhalten, ohne Rücksicht auf dessen Konsequenzen, wird es in der Realität vermutlich genauso selten geben, wie ein Verhalten das einzig auf einer Kosten-Nutzen-Betrachtung beruht, ohne moralische Regeln zu berücksichtigen.

Warum auch „gute“ Menschen nicht immer das tun, was moralisch richtig ist erklären Kohlberg und Candee (1984/1995)<sup>2</sup> wie folgt. Sie unterscheiden zwei Urteile, die sie für das moralische Handeln für notwendig erachten. Zum einen das deontische Urteil, das auf einer übergeordneten Regel oder einem Prinzip basiert, welche Handlung (moralisch) richtig ist. Zum anderen das Urteil über die Verantwortlichkeit, das, was man für moralisch richtig hält, auch in eine konkrete Handlung überführen zu müssen. Für gewöhnlich geht dem Verantwortlichkeitsurteil ein deontisches Urteil voraus. Kohlberg und Candee betonen den starken Einfluss der Situation und verweisen auf mögliche Diskrepanzen zwischen moralischem Urteil (kognitiv) und der Handlung. Menschen wissen manchmal genau, was moralisch richtig ist, und handeln dennoch anders, z.B. weil situative Einflüsse (z.B. die günstige Gelegenheit unbemerkt zu mogeln) sie dazu verleiten, bzw. weil das Verantwortlichkeitsurteil keinen Handlungsbedarf für das Individuum ergibt.

Bei einer Bestrafung werden meist beide Ziele eine Rolle spielen (Mackie, 1985), insbesondere bei der Gegenüberstellung ökonomischer und psychologischer Betrachtungen sozialer Dilemmasituationen ist die Differenzierung aber durchaus sinnvoll. In der Diskussion um retributive Gerechtigkeit wird verschiedentlich auf „disinterested retribution“ (Vidmar, 2001, S. 47 ff., siehe oben das Zitat von Westermarck zur „disinterestedness“) eingegangen, aus dem sich das Konzept des Third-Party-Punishment ableiten lässt. Es geht darum, dass eine unbeteiligte, mit dem Opfer nicht verwandte dritte Person, aufgrund ihrer moralischen Emotionen, strafend eingreift (Heider, 1958). Erklärt wird dieses Vergeltungsverhalten damit, dass durch das Unrecht eine (gesellschaftliche) Norm verletzt wird, an die sich die bestrafende Person gebunden fühlt. Vidmar (2001, S. 56f) verweist darauf, dass beiden Formen der Bestrafung ähnliche Mechanismen zugrunde liegen: „Additionally, attention has been drawn to the

---

<sup>2</sup> Die Frage, ob Moralerwerb kognitiv-entwicklungsorientiert erklärt werden sollte oder über Sozialisationsprozesse wird hier nicht weiter behandelt (Kohlberg, 1976/1995).

fact that 'disinterested' punishment responses [...] are not essentially different from those of the immediate victims of a harm [...]. The common threads are the moral values that arise out of socialization and group membership, broadly conceived. At the psychological level retributive motives arising among persons who are the direct targets of a harm, are different from 'disinterested' observers primarily in the personalization of the harm."

Wie bei altruistischer Bestrafung ist auch bei Rache der Auslöser ein entstandener (subjektiver) Schaden. Gollwitzer (2004) sieht zudem ein Ziel der Rache darin, für Gerechtigkeit zu sorgen. Bestrafung und Rache als Akte retributiver Gerechtigkeit lassen sich nur schwer voneinander trennen. Gollwitzer (2004) verweist darauf, dass sowohl Rache als auch Bestrafung retributive Reaktionen sind, denen die gleichen psychologischen Mechanismen – Ausgleich und Wiederherstellung – zugrunde liegen (vgl. auch Vidmar, 2001; Heider, 1958).

Zusammenfassend lässt sich also festhalten, dass die distributiven Gerechtigkeitsprinzipien in sozialen Dilemmasituationen die Regeln darstellen, nach denen Erträge und Kosten aufgeteilt werden sollten. Wird gegen diese Regeln verstoßen, reagieren sowohl die direkt Betroffenen als auch Nicht-Betroffene, die sich aber an diese Regeln gebunden fühlen, verärgert und bestrafen den Normbrecher. Dabei können ökonomische Kosten-Nutzen-Erwägungen eine Rolle spielen, müssen es aber nicht.

### **3.1.2. Soziale Normen**

Normen sind Regeln, die das menschliche Zusammenleben steuern. Allgemein wird unter einer sozialen Norm eine Regel verstanden, über die ein gewisser Grad an Einigkeit besteht und die mit Hilfe sozialer Sanktionen durchgesetzt wird (Elster, 1990; Horne, 2004). Elster (1990, S. 864f) verweist darauf, dass Verhaltensvorschriften, die aus Normen abgeleitet werden können, nicht an den Verhaltens-Konsequenzen orientiert sind, sondern an höheren Prinzipien: „Social norms are nonconsequentialist obligations and interdictions [...]. Still



further theories, like Kant's moral philosophy, rest on nonconsequentialist obligations. These are quite similar to social norms" (vgl. die Ausführungen weiter oben zu deontischen Gerechtigkeitsauffassungen).

Normen werden im Laufe sozialer Lernprozesse internalisiert (Sozialisation), indem normkonformes Verhalten durch relevante Sozialisationsagenten (Eltern, Freunde, Gesellschaft) verstärkt wird. Für das Einhalten dieser internalisierten Normen bedarf es dann keiner externen Belohnung oder Bestrafung mehr. Diese Prozesse laufen innerhalb der Person ab (Fischer & Wiswede, 2002). Eine Verletzung internalisierter Normen führt zu Schuld- und Schamgefühlen. Sind Normen dagegen nicht internalisiert, müssen sie durch externe soziale Kontrolle und Sanktionen durchgesetzt werden (Wiswede, 1998). Wiswede (2001, S. 341) nennt Normen „stabile Bollwerke“ gegen einen Verfall der Solidarität und verweist darauf, dass Solidarität auch dann weiter besteht, wenn externe Kontrollen entfallen.

Die stärkste Form der Internalisierung lässt soziale Norm und Eigeninteresse verschmelzen. Das Individuum möchte sich an die Norm halten und es ist dann sein (Eigen-)Interesse, sich an die (Gruppen-) Norm zu halten. Im Allgemeinen ist jedoch nicht davon auszugehen, dass ein hundertprozentiger Zusammenhang zwischen der Norm und einem beobachtbaren Verhalten besteht (Horne, 2003).

Das Konzept der Norminternalisierung und der damit verbundenen Internalisierung der Sanktionsinstanz, beinhaltet eine Verbindung zu (moralischen) Emotionen (z.B. Scham, Schuld, Ärger), auf die weiter unten genauer eingegangen wird. Danach entstehen moralische Emotionen nur, wenn eine internalisierte Norm verletzt wird. Das erklärt auch, warum moralische Emotionen eine notwendige Voraussetzung für die Bestrafung sind. Warum sollte eine Person, die eine Verteilungsnorm nicht internalisiert hat, und die daher bei deren Verletzung durch eine andere Person auch keinen Ärger empfindet, die Kosten für die Bestrafung dieser Person auf sich nehmen?

Die Antwort sehen verschiedene Autoren in Meta-Normen, die die Belohnung derjenigen vorschreiben, die die Normeinhaltung durch Bestrafung durchsetzen (Horne, 2001). Heckathorn (1988, S. 540) bietet eine andere Erklärung für die Einhaltung und Durchsetzung von Normen in einer Gesellschaft an. Er postuliert so genannte „regulatory interests“, also ein Bedürfnis, das Verhalten anderer zu kontrollieren, das er aus grundlegenden Präferenzen (z.B. Präferenz für physische Unversehrtheit) und der Bedrohung durch kollektive Bestrafung ableitet.

Wiswede (1998, S. 156 f.) verweist darauf, dass es durch „Kooperationsmechanismen“ zur Auflösung einer sozialen Dilemmasituation kommen kann, so dass ein für alle zufrieden stellendes Ergebnis möglich wird. Hierzu zählt er unter anderem soziale Normen, die Kooperation vorschreiben und implizite Verpflichtungsregeln, die auch Vorstellungen über distributive Gerechtigkeit einschließen.

### **3.1.3. Persönlichkeit und Altruistische Bestrafung: Die Ungerechtigkeitssensibilität**

Bei der persönlichkeitspsychologischen Betrachtung von Gerechtigkeit geht es unter anderem darum festzustellen, wie stabile Persönlichkeitseigenschaften Gerechtigkeitsurteile und Reaktionen auf (Un-)Gerechtigkeit beeinflussen.

Unter einer (Persönlichkeits-)Eigenschaft (trait) soll Ajzen (1988, S. 7) folgend eine stabile, dauerhafte und allgemeine Disposition verstanden werden. Eine solche Persönlichkeitseigenschaft ist situationsübergreifend und bezieht sich nicht auf ein bestimmtes externes Ziel oder Objekt. Da sie auf das Individuum selbst fokussiert, eignet sie sich zur Differenzierung zwischen Personen in der gleichen Situation (z.B. in der Rolle des Aufteilenden im Diktatorspiel) und zur Beschreibung von Persönlichkeitstypen. Sie drückt eine Tendenz zu bestimmten Reaktionen aus („response tendencies“) und kann eine wertende Komponente enthalten. Allgemeine Eigenschaften haben

einen Einfluss auf enger definierte Verhaltenstendenzen und sie wirken auf diese Weise auch auf das Verhalten einer Person<sup>3</sup>.

Betrachtet man also verschiedene Personen in der gleichen Situation, kann man manchmal erhebliche Verhaltensunterschiede feststellen. Diese Unterschiede lassen sich mit den Eigenschaften dieser Personen bzw. der Interaktion der Eigenschaften mit den Charakteristika der Situation erklären.

### *Person oder Situation?*

Über die Frage, was das Verhalten von Menschen beeinflusst – die Persönlichkeit oder die Situation –, wurde in den 60er bis 80er Jahren des 20. Jahrhunderts heftig diskutiert. Die extremen Standpunkte (nur die Person spielt eine Rolle vs. nur die Situation beeinflusst das Verhalten) wurden eigentlich von niemandem wirklich vertreten und die Debatte löste sich zugunsten eines vermittelnden Person-Situation-Standpunktes auf. Hier wird beiden Variablen ein Einfluss zugebilligt. Würden nämlich alle Menschen die Situation gleich wahrnehmen und entsprechend handeln (rein situationistisch), ließe sich keine Variabilität im Verhalten verschiedener Versuchspersonen in Laborexperimenten finden. Diese Variabilität ist aber unbestritten vorhanden. Offensichtlich werden die Situationen also nicht von allen Menschen gleich wahrgenommen bzw. interpretiert. Hagen und Hammerstein (2006) verweisen darauf, dass dann, wenn kein explizites „Framing“ der Situation durch den Versuchsleiter bzw. die Instruktionen vorgenommen wird, die Versuchspersonen selbst einen passenden Rahmen („frame“) konstruieren. Sie verweisen bei der Konstruktion des Rahmens auf die sozialen Hinweisreize, die die Versuchspersonen für die Konstruktion gebrauchen. Welche Reize die Versuchspersonen nutzen und wie sie daraus den Rahmen bei fehlender Information und konstanten, kontrollierten Umgebungsvariablen konstruieren, dürfte maßgeblich von der Persönlichkeit der

---

<sup>3</sup> AJZEN (1988, S.20): „In short, the links from traits to behavior proceed from general personality characteristics to more narrowly defined behavioral tendencies which, in turn, result in relatively specific response dispositions.“

Versuchsperson abhängen. Die Persönlichkeit wirkt sich dementsprechend auf das Framing aus, was seinerseits wiederum die Aktivierung bestimmter Schemata beeinflusst. Zusammen mit dem Framing bedingt das aktive Schema auch die Emotionen und das Verhalten der Versuchsperson in der Situation.

Bei den weiter unten geschilderten experimentellen Spielen handelt es sich eher um „schwache“ Situationen (Mischel, 1976), die weniger klar sind und eine Vorhersage des Verhaltens einer Person allein aus der Situation heraus erschweren. Bei der Analyse des Verhaltens muss dann die Persönlichkeit berücksichtigt werden, da schwache Situationen der individuellen Strukturierung bedürfen und möglicherweise Unsicherheit über das in dieser Situation angebrachte Verhalten herrscht. Obwohl Situationen also erheblichen Einfluss auf das Verhalten haben, sind Persönlichkeitsvariablen zur Vorhersage unverzichtbar. Will man außerdem Vorhersagen über das Verhalten verschiedener Personen in der gleichen Situation machen, kann das nur über differenzierende Persönlichkeitsvariablen erfolgen.

Im Folgenden wird die stabile Persönlichkeitseigenschaft „Ungerechtigkeitssensibilität“ beschrieben (Schmitt, Gollwitzer, Maes & Arbach, 2005), die in der Vergangenheit bereits mehrfach erfolgreich zur Analyse von Reaktionen auf ungerechtes Verhalten eingesetzt wurde.

#### *Ungerechtigkeitssensibilität*

Das Konzept der Ungerechtigkeitssensibilität (im Folgenden auch abgekürzt mit US) sagt, dass Menschen sich in Bezug auf ihre Toleranz gegenüber Ungerechtigkeit und moralischem Normbruch unterscheiden (Schmitt et al., 2005).

Grundsätzlich kann eine Person Ungerechtigkeit aus drei Perspektiven wahrnehmen (Schmitt et al., 2005; Mikula, 1986):

1. Als Opfer: Eine Person erleidet durch das Handeln einer anderen Person einen (ungerechten) Schaden.

2. Als Nutznießer / Täter: Eine Person profitiert vom (eigenen) Handeln, das einer anderen Person einen (ungerechten) Schaden zufügt.
3. Als Beobachter: Eine Person beobachtet, wie eine andere Person durch das Handeln einer dritten Person einen (ungerechten) Schaden erleidet.

Aus diesem Grund besteht das Erhebungsinstrument der US auch aus drei Teilfragebögen (Schmitt et al., 2005). Je nach Perspektive wird die Wahrnehmung und Bewertung einer Ungerechtigkeit sowie die Reaktion darauf unterschiedlich ausfallen. Ein und dieselbe Person kann unterschiedlich stark auf Ungerechtigkeiten reagieren, je nachdem, ob sie selbst Opfer (im Folgenden mit  $US_{\text{Opfer}}$  abgekürzt), Beobachter ( $US_{\text{Beobachter}}$ ) oder Nutznießer / Täter ( $US_{\text{Nutznießer}}$ ) ist. Es gibt z.B. Personen, die sehr sensibel auf Ungerechtigkeit reagieren, wenn sie selbst Opfer (hohe Werte auf  $US_{\text{Opfer}}$ ) sind, aber relativ unberührt bleiben, wenn das gleiche Schicksal andere trifft (niedrige Werte auf  $US_{\text{Beobachter}}$  bzw.  $US_{\text{Nutznießer}}$ ). Auf der anderen Seite gibt es Personen, die starke Schuldgefühle entwickeln, wenn sie (ungerecht) einen Vorteil erlangen (hohe Werte auf  $US_{\text{Nutznießer}}$ ), aber nicht so stark auf ungerechte Benachteiligungen gegen sich selbst reagieren (niedrige Werte auf  $US_{\text{Opfer}}$ ).

Aus diesen Beispielen lassen sich schon verschiedene Verbindungen zu anderen Persönlichkeitseigenschaften erahnen. Schmitt et al. (2005) konnten zeigen, dass  $US_{\text{Opfer}}$  positiv mit verschiedenen antisozialen Persönlichkeitseigenschaften korreliert (z.B. Machiavellismus, Rache, Eifersucht, Neurotizismus) und negativ mit prosozialen (z.B. Vertrauen, Verträglichkeit). Auf der anderen Seite korrelieren  $US_{\text{Beobachter}}$  und  $US_{\text{Nutznießer}}$  positiv mit prosozialen Eigenschaften (z.B. Fähigkeit zur Rollenübernahme, Empathie, soziale Verantwortung, Verträglichkeit). Aus den Korrelationsmustern lässt sich ableiten, dass  $US_{\text{Beobachter}}$  und  $US_{\text{Nutznießer}}$  Gerechtigkeitsbelange in einer klareren Form widerspiegeln, während die Opfersensibilität neben so-

zialen Gerechtigkeitsanliegen in starkem Maße auch von Selbstschutz und sogar egoistischen Motiven getrieben wird. Die inhaltliche Nähe von  $US_{\text{Beobachter}}$  zu  $US_{\text{Nutznießer}}$  findet sich auch in den Korrelationsmustern der drei Dimensionen wieder; sie sind nämlich nicht unabhängig voneinander. Während  $US_{\text{Beobachter}}$  und  $US_{\text{Nutznießer}}$  recht hoch miteinander korrelieren, sind die Korrelationen zwischen  $US_{\text{Beobachter}}$  und  $US_{\text{Opfer}}$  durchweg niedriger. Zwischen  $US_{\text{Nutznießer}}$  und  $US_{\text{Opfer}}$  lassen sich für gewöhnlich nur sehr niedrige oder gar keine signifikanten Korrelationen finden (Schmitt et al., 2005; Fetchenhauer & Huang, 2004).

Ungerechtigkeitssensibilität lässt sich an vier Indikatoren festmachen (Schmitt, Neumann & Montada, 1995):

1. Häufigkeit, mit der Ungerechtigkeit wahrgenommen wird;
2. Intensität des Ärgers, der als Reaktion auf die Ungerechtigkeit entsteht;
3. Intrusivität, die Aufdringlichkeit, mit der Gedanken an die Ungerechtigkeit auftreten;
4. Punitivität, die Stärke des Wunsches, den Täter zu bestrafen

Die Ungerechtigkeitssensibilität wird über eine Skala mit 30 Items gemessen (Schmitt, Maes & Schmal, 1997). Je 10 Items repräsentieren  $US_{\text{Opfer}}$ ,  $US_{\text{Beobachter}}$  und  $US_{\text{Nutznießer}}$ . Die drei Subskalen unterscheiden sich durch die abgefragten Reaktionen auf Ungerechtigkeit und durch den Text der Instruktionen. Ein Item aus der Opferperspektive lautet zum Beispiel: "Es ärgert mich, wenn es anderen unverdient besser geht als mir", während das analoge Item aus der Beobachterperspektive lautet: "Ich bin empört, wenn es jemandem unverdient schlechter geht als anderen" und aus der Nutznießer- / Täterperspektive: "Ich habe Schuldgefühle, wenn es mir unverdient besser geht als anderen" (Alle drei Subskalen befinden sich im Anhang dieser Arbeit bzw. bei Schmitt et al., 2005; Schmitt et al., 1997).

Verschiedene Studien ergaben, dass die US in der Lage ist, Reaktionen in ungerechten Situationen zu erklären bzw. vorherzusagen.

Schmitt und Mohiyeddini (1996) konnten zeigen, dass sich die Gerechtigkeitsbeurteilung einer Lotterie zur Vergabe knapper Seminarplätze und die Zustimmung zu Aktivitäten des Studentenparlaments gegen diese Lotterie durch  $US_{\text{Opfer}}$  voraussagen lassen<sup>4</sup>.

Gollwitzer et al. (2005) konnten in einer Reihe von Untersuchungen nachweisen, dass  $US_{\text{Opfer}}$  entweder negativ oder gar nicht mit prosozialem Verhalten zusammenhängt, wohingegen sich für  $US_{\text{Nutznießer}}$  ein positiver Zusammenhang zeigt. Für verschiedene Verhaltensweisen (z.B. Ladendiebstahl, Schwarzfahren, Schwarzarbeit) zeigte sich, dass Versuchspersonen mit hohen Werten auf  $US_{\text{Opfer}}$  eher bereit waren, gesellschaftliche Normen zu brechen, wenn es sich lohnte. Im Gegensatz dazu hielten sich Versuchspersonen mit hohen Werten auf  $US_{\text{Nutznießer}}$  tendenziell eher an die Normen.

Fetchenhauer und Huang (2004) untersuchten den Einfluss der Ungerechtigkeitssensibilität in experimentellen Spielen. Sie fanden einen Zusammenhang von US und der Aufteilung von Geldbeträgen bzw. der Ablehnung von Angeboten. Die Versuchspersonen mit hohen Werten auf  $US_{\text{Opfer}}$  tendierten dazu, eigennützig und strategisch zu handeln, während Versuchspersonen mit hohen Werten auf  $US_{\text{Nutznießer}}$  die Geldbeträge eher fair verteilten.

Betrachtet man neben den eben geschilderten empirischen Ergebnissen die Korrelationsmuster mit anderen Persönlichkeitsdimensionen, ergibt sich ein recht eindeutiges Bild (Gollwitzer et al., 2005):

- Hohe Werte auf  $US_{\text{Opfer}}$  repräsentieren Eigeninteresse und den Willen zu strategischem Verhalten. Hierbei wird evtl. auch Normbruch und Schaden für andere in Kauf genommen (z.B. Handlungen unterlassen, die schlechter Gestellten helfen könnten).

---

<sup>4</sup> In dieser Untersuchung wurde allerdings noch eine frühere längere Skala (SBI = Sensitivity to Befallen Injustice) verwendet.

- Hohe Werte auf  $US_{\text{Nutznießer}}$  sind verbunden mit prosozialen Einstellungen und Verhaltensmustern. Diese Menschen haben ein echtes Interesse an anderen Menschen und Gerechtigkeit.
- $US_{\text{Beobachter}}$  liegt von den Reaktionen näher an  $US_{\text{Nutznießer}}$ , weswegen manchmal eine gemeinsame Auswertung für diese beiden Dimensionen zusammen vorgenommen wird (bei Fetchenhauer & Huang, 2004, erfolgt die Auswertung sowohl einzeln als auch gemeinsam). Allerdings bestehen eben doch auch substantielle Korrelationen mit  $US_{\text{Opfer}}$ . Aufgrund empirischer und theoretisch-inhaltlicher Überlegungen werden in der Untersuchung von Gollwitzer et al. nur  $US_{\text{Nutznießer}}$  und  $US_{\text{Opfer}}$  betrachtet.

Zur Vorhersage von pro- bzw. antisozialem Verhalten scheinen sich die Dimensionen  $US_{\text{Nutznießer}}$  und  $US_{\text{Opfer}}$  besser zu eignen, als  $US_{\text{Beobachter}}$ . Theoretisch-inhaltlich liegt dies nahe, und die empirischen Ergebnisse bestätigen dies auch weitgehend.

### 3.1.4. Emotionen<sup>5</sup>

[...] and the idea of their distress serves only to inflame and blow up our animosity against those who occasioned it. When we think of the anguish of the sufferers, we take part with them more earnestly against their oppressors; we enter with more eagerness into all their schemes of vengeance, and feel ourselves every moment wreaking, in imagination, upon such violators of the laws of society, that punishment which our sympathetic indignation tells us is due to their crimes.

(Smith, 1759/1982)

Auf die enge Verbindung zwischen erlebter Ungerechtigkeit und Emotionen wurde bereits verschiedentlich in dieser Arbeit verwiesen. Da sie bei der Equity-Theorie, bei der Reaktion auf Normbruch und bei der Ungerechtigkeitssensibilität zentral sind, wird hier genauer auf die Rolle der Emotionen eingegangen.

---

<sup>5</sup> Im Folgenden werden die Begriffe „Emotion“ und „Gefühl“ synonym verwendet.



Emotionen entstehen als Reaktion auf bestimmte Situationen. Sie sind subjektive Erfahrungen, in deren Mittelpunkt eine Valenzbewertung (angenehm/unangenehm) sowie die Erregung steht, die mit ihnen einhergeht. Emotionen versetzen ein Individuum in einen Zustand der Handlungsbereitschaft („state of action readiness“) und bereiten auf diese Weise bestimmte Handlungen vor (Frijda, 1988, S. 351) bzw. motivieren sie. Frank (1988) nennt Gefühle und Emotionen proximate Ursachen des Verhaltens. Fetschenhauer und Bierhoff (2004, S. 135) bezeichnen sie als „Motor“, der Reziprozität in sozialen Austauschbeziehungen antreibt“. Für den Bereich der Risikobewertung und Reaktion auf Risiken konnten Loewenstein, Weber, Hsee und Welch (2001) zeigen, dass bei Abweichungen zwischen der kognitiven Bewertung und der emotionalen Reaktion auf Risiken häufig die emotionalen Reaktionen das Verhalten determinieren.

Emotionen können auf verschiedene Weise gemessen werden. Eine einfache und ökonomische Art ist die Selbstauskunft über einen Fragebogen. Andere Verfahren sind die physiologische Messung (z.B. Messung des Hautwiderstandes) oder die Feststellung der Aktivität von Hirnarealen durch bildgebende Verfahren der Hirnforschung, auf die weiter unten, im Abschnitt über die neuroökonomischen Befunde, eingegangen wird.

Verschiedene Autoren verweisen darauf, dass starke Emotionen kognitive Prozesse zurückdrängen bzw. die Komponenten der Bewertung von Kosten und Nutzen verändern können (vgl. z.B. Pillutla & Murnighan, 1996; Frank, 1988; Kahneman, 2003). So z. B. Frijda (1988, S. 355): „[...] emotions know no probabilities. They do not weigh likelihoods“. Greene und Haidt (2002; vgl. auch Haidt, 2001) betonen den großen Einfluss von Emotionen und affektiven Intuitionen auf moralische Urteile. Kognitive Abwägungen spielen ihrer Meinung nach eine signifikante, aber geringere Rolle. Hieraus sollte allerdings nicht leichtfertig ein Widerspruch zwischen Ratio und

Emotion abgeleitet werden; beide sind vermutlich immer nötig, um gute Urteile zu fällen (Damasio, 1994; Frank, 1988; Frijda, 1988).

Kosten-Nutzen-Analysen werden bei starken Emotionen also entweder nicht mehr durchgeführt oder die Gewichte der einzelnen Kosten- und Nutzenkomponenten werden gegenüber dem Normalzustand ohne starke Emotionen dramatisch verändert (Frank, 1988). Emotionen können zu Handlungen führen, deren Konsequenzen wiederum Emotionen anderer Valenz auslösen.

*Emotionen als Reaktion auf ungerechte Erlebnisse und als Auslöser von Bestrafung*

Mit der Wahrnehmung einer Situation, in der Normen der Fairness verletzt werden, sind konkrete Emotionen verknüpft. Mikula, Scherer und Athenstaedt (1998) berichten, dass die häufigsten emotionalen Reaktionen auf Ungerechtigkeit Ärger und Empörung sind (vgl. auch Mikula, 1986; Hassebrauck, 1984). Ärger ist eine stark negative Emotion, die mit hoher Erregung und dem Bedürfnis einhergeht, etwas gegen die den Ärger auslösenden Umstände zu tun (Frijda, 1988). Eine Möglichkeit, etwas gegen die Ursache des Ärgers zu unternehmen und ihn abzubauen, ist die Bestrafung des Täters (de Quervain et al., 2004; Elster, 1998). Elster (1994) nennt als Handlungstendenz, die auf Ärger folgt, den Wunsch, die Person, die einen geschädigt hat, ihrerseits zu schädigen. Verschiedene Autoren sehen starke negative Emotionen (Ärger, Wut, Empörung, Neid usw.) als Ursache für die Bestrafung (z.B. Pillutla & Murnighan, 1996; Elster, 1998; Brandstätter & Königstein, 2001; Fehr & Gächter, 2002; Fehr & Fischbacher, 2004b; Skarlicki & Kulik, 2005; Xiao & Houser, 2005; Masclet & Villeval, 2006; van't Wout, Kahn, Sanfey & Aleman, 2006). Hassebrauck (1984, S. 83) nennt „[...] emotionale Konsequenzen distributiver Ungerechtigkeit eine notwendige Bedingung für das Wiederherstellen von Ausgewogenheit[...]“ (vgl. auch Austin & Walster, 1974; im Gegensatz dazu, Kohlberg & Candee, 1984/1995, S. 379: "Wir haben deshalb den Eindruck, dass emotionale Beteili-

gung keine notwendige psychische Determinante für die Definition moralischen Verhaltens darstellt."). Pillutla und Murnighan (1996) gehen davon aus, dass die reine Wahrnehmung der Ungerechtigkeit für eine Bestrafung nicht ausreicht. Ihrer Meinung nach muss immer auch Ärger vorliegen, damit eine Person unter eigenen Kosten bestraft (vgl. auch Miller, 2001).

Die Bestrafung wiederum kann zu positiven Emotionen führen, die für das Individuum belohnend wirken (de Quervain et al., 2004; Bosman & van Winden, 2002). Materieller Schaden wird auf diese Weise evtl. emotional ausgeglichen (Bosman & van Winden, 2002, sprechen von einem trade-off zwischen emotionaler Befriedigung durch die Bestrafung und materiellen Erträgen; ebenso Bosman, Sutter & van Winden, 2005).

Haidt (2001) geht davon aus, dass moralische Urteile meist schnell und automatisch ausgelöst werden. Moralische Urteile, die schnell und emotional, ohne große Beteiligung kognitiver Prozesse ablaufen, wären in der Diktion Kahnemans (2003) dem System 1 zuzuordnen. Kahneman unterscheidet das schnelle, emotional gefärbte, automatische System 1, über das die Wahrnehmung und die Intuition ablaufen, von System 2, das für die kontrollierte, neutrale kognitive Verarbeitung zuständig ist und dadurch deutlich langsamer arbeitet. Vermutlich sind beide Systeme bei Gerechtigkeitsurteilen involviert; manche Erkenntnisse (vgl. z.B. Greene & Haidt, 2002) legen aber nahe, dass System 1 die auf Ungerechtigkeit folgenden Handlungen steuert.

Camerer (2003; vgl. auch Smith, 1759/1982) weist zu Recht darauf hin, dass zwischen den Emotionen, die ein direkt von Ungerechtigkeit Betroffener empfindet (Ärger), und denen, die eine unbeteiligte dritte Partei empfindet (Empörung), unterschieden werden muss (darauf, dass verschiedene Situationen zu verschiedenen Emotionen führen verweist auch Frijda, 1988). Camerer vermutet, dass Empörung eine geringere Intensität aufweist als Ärger und dass dies die Ursache für

die schwächere Bestrafung durch unbeteiligte Dritte ist. Dieses Argument ist durchaus schlüssig, allerdings stellt sich die Frage, ob Menschen, wenn sie nach ihren Gefühlen befragt werden oder wenn sie die Gefühle anderer bewerten, diese Unterscheidung auch vornehmen (können). Ganz allgemein kann allerdings gesagt werden, dass ein starker Zusammenhang der Wahrnehmung von Ungerechtigkeit und verschiedenen negativen Gefühlen besteht (Mikula et al., 1998). Eine genauere Differenzierung der Emotionen, die bei selbst erlebter und bei beobachteter Ungerechtigkeit entstehen, wird weiter unten bei den Überlegungen zur Ungerechtigkeitssensibilität vorgenommen.

Auf einen weiteren Aspekt der Emotionen bei Bestrafungen weisen Hopfensitz und Reuben (2005) hin. Damit die Bestrafung eine positive Wirkung entfalten kann, ist es notwendig, dass der Bestrafte sich in sein Schicksal fügt und sich nicht seinerseits rächt. Rache könnte schnell zur Vernichtung einer großen Menge von Ressourcen führen, was für alle Beteiligten schädlich wäre. Dazu sollte es aber nicht kommen, wenn der Bestrafte sich schuldig fühlt und die Strafe annimmt. Ein Beispiel für negative Folgen von Bestrafung und Gegenbestrafung liefert Nikiforakis (2004; vgl. auch die Ausführungen zu Blutfehden bei Elster, 1990)<sup>6</sup>.

### *Neuro-Ökonomische und physiologische Erkenntnisse*

In den letzten Jahren wurde zur Erklärung des Verhaltens in experimentellen Spielen auch auf (bildgebende) Verfahren der Hirnforschung (z.B. funktionelle Magnetresonanztomographie, functional magnetic resonance imaging; fMRI) zurückgegriffen. Mit Hilfe dieser Verfahren lässt sich die Aktivität verschiedener Hirnareale sichtbar machen. So kann auf der Basis der Gehirnfunktionen die Rolle der Emotionen und Kognitionen bei ökonomischen Entscheidungen

---

<sup>6</sup> Er nennt in den Schlussfolgerungen negative Emotionen als mögliche Begründung für die Gegenbestrafung, ohne weiter darauf einzugehen. Eine Untersuchung und genaue Differenzierung insbesondere vor dem Hintergrund der Ergebnisse von Hopfensitz und Reuben (2005) wäre wünschenswert.

untersucht werden (z.B. Sanfey, Rilling, Aronson, Nystrom & Cohen, 2003).

De Quervain et al. (2004) finden Hinweise darauf, dass es belohnend sein kann, einen unfairen Interaktionspartner zu bestrafen. Bei Versuchspersonen, die eine starke Bestrafungsneigung aufwiesen, zeigte sich eine verstärkte Aktivierung von Hirnregionen, die mit der Verarbeitung von Belohnungen assoziiert sind. Außerdem fanden die Autoren folgenden Zusammenhang: Personen, die eine hohe Aktivierung dieser Hirnregionen aufwiesen, waren auch bereit, höhere Kosten für die Bestrafung einzugehen. Dieses Ergebnis wird so interpretiert, dass die Aktivierung die Antizipation der Genußtuung über die Bestrafung widerspiegelt.

### **3.1.5. Ökonomische Modelle: Soziale Präferenzen**

Das Modell des rational eigennützigen Homo Oeconomicus kann altruistische Bestrafung nicht erklären, denn sie führt im (materiellen) ökonomischen Sinne nicht zu optimalen Ergebnissen für ein Individuum und sollte daher aus dieser Sichtweise nicht vorkommen. Diese ausschließliche Konzentration auf die eigene (materielle) Nutzenmaximierung ist offensichtlich eine zu beschränkte Sichtweise.

Innerhalb eines erweiterten ökonomischen Erklärungsrahmens werden zur Erklärung reziproken und altruistischen Verhaltens soziale Präferenzen einbezogen. Die Akteure in diesen Modellen berücksichtigen neben ihrem eigenen Nutzen auch das Wohlergehen anderer Menschen. Menschen haben demnach ein Interesse daran, wie z.B. eine bestimmte Summe Geld aufgeteilt werden soll, und erwarten nicht, alles selbst zu bekommen (Fehr & Fischbacher, 2002).

#### *Ungleichheitsaversion*

Verschiedene Modelle versuchen, die Ergebnisse in experimentellen Spielen mit Hilfe der absoluten und relativen Erträge aus dem Spiel zu erklären (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999).

Demnach reagieren Menschen ablehnend auf Ergebnisse, wenn diese ungleich sind.

Innerhalb der verschiedenen Modelle werden unterschiedliche Vergleichsstandards angenommen: Je nach Modell vergleichen die Spieler sich entweder mit allen anderen (Fehr & Schmidt, 1999) oder mit dem Gruppendurchschnitt (Bolton & Ockenfels, 2000) und achten bei diesen Vergleichen auf Ungleichheit. Sie reagieren in diesen Modellen allerdings nur dann auf Ungleichheit negativ, wenn sie selbst betroffen sind, nicht aber auf mögliche Ungleichheiten, die zwischen den anderen Spielern bestehen. Dies steht im Widerspruch zu den weiter oben gemachten Aussagen über disinterested punishment und der Tatsache, dass internalisierte Verteilungsnormen u. U. auch dann durchgesetzt werden, wenn eine Person nicht direkt vom Normbruch betroffen ist.

Nach Fehr und Schmidt (1999) nehmen Menschen Ungleichheit (negativ) wahr, sowohl wenn sie besser als auch wenn sie schlechter als andere gestellt sind. Sie leiden allerdings mehr unter einer Schlechterstellung als unter einer Besserstellung. Bei Bolton und Ockenfels (2000) sind beide Abweichungen gleich gewichtet. Die Gleichgewichtung positiver und negativer Abweichungen von einem Vergleichsstandard steht sowohl im Widerspruch zur Equity-Theorie als auch zur Prospect Theorie (Kahneman & Tversky, 1979), die beide eine stärkere Reaktion auf negative Abweichungen annehmen als auf positive.

Beide Modelle, sowohl das von Fehr und Schmidt als auch das von Bolton und Ockenfels, berücksichtigen lediglich die absoluten und die relativen Erträge der Spieler. Die Intentionen, die hinter den Aufteilungen stehen, werden nicht berücksichtigt. Das ist ein großer Schwachpunkt dieser Modelle, denn es wurde vielfach gezeigt, dass die Intentionen eines Normbrechers durchaus eine Rolle für die Reaktionen anderer auf den Normbruch spielen (Mikula, 1993). Kann bei einem Normbrecher Absicht angenommen werden, wird er meist

härter bestraft als wenn keine Intention unterstellt werden kann (vgl. z.B. Miller & Vidmar, 1981; Falk, Fehr & Fischbacher, 2003).

### *Reziprozität*

Diesen Widerspruch wollen Modelle zur Reziprozität auflösen, indem sie neben der Betrachtung der Ergebnisse auch die Intention des Aufteilenden berücksichtigen. Im Modell von Falk und Fischbacher (2006), das hier stellvertretend für andere Modelle vorgestellt werden soll, wird unter einer reziproken Handlung eine gleichgerichtete Reaktion auf ein freundliches oder unfreundliches Verhalten einer anderen Person verstanden. Die Bewertung, ob ein Verhalten freundlich oder unfreundlich ist, hängt von den Ergebnissen der Handlung und von der der Handlung zugrunde liegenden Intention ab. Die Intention wiederum lässt sich aus den Handlungsalternativen erschließen. (Stehen einer Person bei der Aufteilung von 10 Euro zwischen sich und einer anderen Person die Alternativen 6/4 und 9/1 zur Verfügung und entscheidet sie sich, neun Euro zu behalten und einen Euro an die andere Person abzugeben, kann man daraus auf eine unfaire Intention schließen). Aus dem Modell lässt sich eine Tendenz zu reziprokem Verhalten ableiten, die zu einem Teil aus der Ungleichheit der Ergebnisse resultiert, zu einem anderen Teil jedoch aus der Intention des Interaktionspartners (vgl. auch Falk et al., 2003; Fehr & Fischbacher, 2002). Im Gegensatz zu den Modellen der Ungleichheitsaversion wird im Modell von Falk und Fischbacher auch dann ein unfreundliches Verhalten mit einem unfreundlichen Verhalten beantwortet, wenn dies nicht die Ungleichheit verringert.

Durch die Integration der Intentionen ist das Modell von Falk und Fischbacher (2006) näher an der psychologischen Realität sozialer Dilemmasituationen als die Modelle von Fehr und Schmidt (1999) und Bolton und Ockenfels (2000). Für alle drei Modelle gilt allerdings, dass der wichtige Einfluss moralischer Emotionen völlig ausgeblendet wird.

### **3.1.6. Zusammenfassung der proximatén Erklärungen**

Wie schon gesagt, lassen sich in der Realität selten Verhaltensweisen finden, die auf den ausschließlichen Einfluss des Eigennutzmotivs hinweisen. Ökonomische Modelle, die dies unterstellen, können also lediglich als normativ-präskriptive Modelle gesehen werden, die wenig deskriptive Erklärungskraft besitzen. Modelle sozialer Präferenzen, die neben dem unbestritten starken Einfluss des Eigennutzstrebens auch soziale Motive, wie Reziprozität oder Ungleichheitsaversion, berücksichtigen, können deutlich mehr erklären.

Ein weiteres Manko ökonomischer Modelle ist, dass die Prozesse, die einer Bestrafung vorangehen, nicht deutlich werden. Hier führt die Betrachtung der psychologischen Prozesse in Verbindung mit sozialen Normen und Gerechtigkeitsprinzipien weiter. Werden diese Regeln verletzt, entstehen starke negative Emotionen, die retributives Verhalten motivieren. Dabei ist insbesondere bemerkenswert, dass Bestrafung nicht dem Eigennutzmotiv folgen muss, diesem manchmal sogar entgegenlaufen kann.

Auch Persönlichkeitsunterschiede führen zu unterschiedlichem Verhalten. Situationen werden von unterschiedlichen Menschen aufgrund ihrer Persönlichkeit unterschiedlich wahrgenommen und sie reagieren auch auf ungerechte Situationen verschieden. Ungerechtigkeitssensible Personen reagieren mit stärkeren negativen Emotionen auf den Bruch von Gerechtigkeitsnormen. Ihr Ärger und ihre Empörung motivieren sie, Gegenmaßnahmen zu ergreifen und den Normbrecher zu bestrafen.

Die Tatsache, dass Emotionen eine gewichtige Rolle für das Verhalten in sozialen Dilemmasituationen spielen, dürfte mittlerweile relativ unstrittig sein. Offen und heiß diskutiert ist allerdings die Frage, wie entsprechende Emotionen sich im Rahmen der Evolution herausbilden konnten und durch welche Mechanismen die negativen materiellen Fitness-Folgen, die durch eine Bestrafung entstehen, ausge-



glichen werden. Diese Fragen werden im Rahmen der ultimatzen evolutionären Erklärungen diskutiert.

### **3.2. Ultimate Erklärungen für Altruistische Bestrafung**

Eine befriedigende ultimate Erklärung muss folgende Frage beantworten können: Wenn altruistische Bestrafung einen Fitnessnachteil für das Individuum mit sich bringt, warum sind dann Individuen, die solche Verhaltensweisen zeigen, nicht im Lauf der Zeit ausgestorben? Die Darwinsche Evolutionstheorie tut sich mit der Erklärung eines Verhaltens schwer, das die Fitness eines Individuums senkt. Außerdem müssen sich aus den ultimatzen Erklärungen die im vorangegangenen Abschnitt erläuterten proximatzen Erklärungen ableiten lassen. Einige Theorien, die in der Lage sind altruistisches Verhalten zu erklären, werden in den nächsten Abschnitten vorgestellt, und ihre Eignung zur Erklärung von altruistischer Bestrafung wird diskutiert.

#### **3.2.1. Verwandtschaftsaltruismus: Blut ist dicker als Wasser**

In der Theorie von Hamilton (1963; 1964) wird davon ausgegangen, dass es Altruismus ausschließlich zwischen Verwandten (kin) geben kann. Durch Altruismus erhöht sich die reproduktive Fitness der verwandten Individuen und Gene, die den eigenen gleichen, werden mit höherer Wahrscheinlichkeit an Nachfahren weiter gegeben. Wenn auf diese Weise mehr Nachkommen und nahe Verwandte des Individuums überleben und sich ihrerseits fortpflanzen können, geht die Rechnung auf. Mit abnehmendem Grad der Verwandtschaft nimmt entsprechend dieser Theorie die Hilfsbereitschaft zwischen Individuen ab. Die Regel, die diesen Zusammenhang mathematisch ausdrückt wird auch Hamilton's Rule genannt:  $b > c/r$  (b: Nutzen für den Empfänger der altruistischen Handlung; c: Kosten für den Altruisten; r: Koeffizient für den Verwandtschaftsgrad).

Dieser Ansatz ist nicht in der Lage, altruistische Handlungen gegenüber Fremden zu erklären (also auch nicht altruistische Bestrafung). Eine geringere reproduktive Fitness auf Grund der Hilfe gegenüber Fremden führt im Rahmen dieses Modells zu weniger Nachkommen und dadurch über viele Generationen zum „Aussterben“ derjenigen Individuen, die eine Tendenz zum Altruismus gegenüber Nicht-Verwandten haben. Offensichtlich stimmt diese Sichtweise nicht mit den empirischen Beobachtungen überein und, obwohl Altruismus gegenüber Blutsverwandten sicherlich eine gewichtige Rolle spielt, muss es weitere Erklärungen geben.

### **3.2.2. Reziproker Altruismus: Wie du mir, so ich dir**

Eine Theorie, die in der Lage ist, Altruismus zwischen Nicht-Verwandten zu erklären, beruht auf der Idee des reziproken Altruismus (Trivers, 1971). Hierbei wird die Interaktionszukunft zwischen den Individuen berücksichtigt, und es ist entscheidend, dass die Kosten des altruistischen Akts für Helfer deutlich geringer sind als der Nutzen derer, denen geholfen wurde. Ein Individuum verhält sich dann altruistisch<sup>7</sup>, wenn es erwarten kann, dass sich sein Interaktionspartner in Zukunft reziprok verhalten wird und bei Gelegenheit ebenfalls altruistisch handelt.

Warum sollte derjenige, dem geholfen wurde, sich reziprok altruistisch verhalten, wenn dies nötig ist? Wenn eine genügend hohe Wahrscheinlichkeit besteht, dass man in Zukunft wieder mit einem bestimmten Partner interagiert (Axelrod & Dion, 1988, S. 1387, sprechen vom "shadow of the future"), und ist der entgangene Nutzen einer nicht erhaltenen altruistischen Hilfe höher als die Kosten für reziprok altruistisches Verhalten, dann haben nicht-altruistische und nicht-reziproke Individuen einen Fitnessnachteil. „[...] under certain conditions natural selection favors these altruistic behaviors because in the long run they benefit the organism performing them.“ (Trivers, 1971, S. 35).

---

<sup>7</sup> Es handelt sich hierbei offensichtlich um Altruismus im biologischen Sinn (s. o.).

Die glaubhafte Androhung, in der Zukunft nicht mehr zu kooperieren, kann reziprokes Verhalten erklären. Reziproker Altruismus beruht nach diesem Verständnis auf einem „Wie-du-mir-so-ich-dir-Prinzip“, das zukünftige Erträge aus einer Interaktion bei der momentanen Entscheidung für oder gegen Kooperation berücksichtigt. Das Funktionieren einer solchen „Tit-for-Tat-Strategie“, bei der auf Kooperation mit Kooperation und auf Nicht-Kooperation mit Nicht-Kooperation reagiert wird, wurde von Axelrod und Hamilton (1981) nachgewiesen.

Reziprok altruistisches Verhalten ist hierbei (langfristig) immer mit positiven Erträgen für das Individuum verbunden, ansonsten werden solche Handlungen unterlassen. Boyd und Richerson (1988) verweisen allerdings kritisch darauf, dass sich Reziprozität vermutlich nur in relativ kleinen Gruppen entwickeln kann. Außerdem ist es erforderlich, dass einer der Interaktionspartner mit einem kooperativen Verhalten beginnt. Schon kleine (zufällige) Störungen des Systems führen zu unkooperativem Verhalten in der Zukunft, wenn Tit-for-Tat stur angewendet wird.

Sehr interessant sind die Äußerungen Trivers (1971, S. 49) über „moralistic aggression“: „Moralistic aggression and indignation in humans was selected for in order [...] to select directly against the unreciprocating individual by injuring, killing or exiling him. [...] Injustice, unfairness, and a lack of reciprocity often motivate human aggression and indignation“. Hier geht Trivers auf die wichtige Rolle von moralischen Emotionen bei altruistischem Verhalten ein und erklärt, wie Selektionsprozesse diese Gefühle und altruistisches Verhalten gefördert haben können. Genau die Gefühle, die Trivers anspricht, wurden im Abschnitt über Emotionen als Auslöser für altruistische Bestrafung identifiziert. Bowles und Gintis (2004) nennen „moralistic aggression“ als motivationale Grundlage von altruistischer Bestrafung.

Fehr, Fischbacher und Gächter (2002) verweisen auf die positive Nutzen-Kosten-Bilanz bei reziprokem Altruismus und benutzen die Wörter „eigennützig“ und „reziprok altruistisch“ synonym. Dies scheint durchaus angebracht, denn das Konzept des reziproken Altruismus hat mit Altruismus – insbesondere im psychologischen Sinne – gar nichts mehr gemein.

Kann die Theorie des reziproken Altruismus erklären, wieso es altruistisches Verhalten in einmaligen Interaktionen gibt? Nein, denn unter solchen Bedingungen existieren keine zukünftigen Erträge, und eine reziprok altruistische Person wird sich dann auch nicht kooperativ verhalten, denn unabhängig davon, wie gering die Kosten für den altruistischen Akt sind, entsteht ein Fitnessnachteil für das Individuum. Eine Theorie, in der Kooperation und Altruismus auf reziprokem Altruismus basieren, ist nicht in der Lage zu erklären, wie sich altruistische Bestrafung – und altruistisches Verhalten allgemein – in einmaligen Interaktionen zwischen Fremden und in großen Gruppen in der Evolution entwickelt hat. Altruistische Bestrafung, insbesondere durch unbeteiligte Dritte, ist in diesem Konzept nicht vorgesehen. Wenn die einzige Möglichkeit, die anderen Mitglieder in der Gruppe zu bestrafen, die Verweigerung des eigenen Beitrags ist, führt schon eine geringe Zahl an unkooperativen Mitgliedern der Gruppe zum Erliegen der Kooperation (Fetchenhauer & Bierhoff, 2004).

### **3.2.3. Indirekte Reziprozität: Wie du ihm, so ich dir**

Evolutionäre Theorien, die auf indirekter Reziprozität fußen, gehen davon aus, dass reziproke Individuen durch ihre altruistischen Taten eine gute Reputation aufbauen können (Nowak & Sigmund, 2005). Die positiven Effekte einer guten Reputation sind, dass man in Zukunft von anderen Menschen auch mit reziprok altruistischem Verhalten rechnen kann bzw. häufig als Interaktions- und Kooperationspartner ausgewählt wird. Diese beidseitig vorteilhaften Interaktionen können die negativen Effekte altruistischen Verhaltens ausgleichen.

Hierbei besteht allerdings das Problem, dass es eine Art soziales Gedächtnis geben muss, in dem die Informationen über vergangenes Verhalten gespeichert werden (Nowak & Sigmund, 2005). Außerdem verhalten sich nach diesem Ansatz Individuen nur dann altruistisch, wenn das dem Aufbau der Reputation dient. Nimmt niemand Notiz von ihrem Verhalten, werden sie ihre eigenen Ressourcen maximieren. Auch in diesem Ansatz wird auf die Wichtigkeit der Evolution moralischer Emotionen verwiesen.

Es muss allerdings kritisch angemerkt werden, dass die Theorien indirekter Reziprozität letztendlich nicht erklären können, warum sich Menschen auch in einmaligen Situationen unter vollständiger Anonymität altruistisch verhalten bzw. warum sie dies in großen Gruppen tun. In vielen Experimenten ist die Möglichkeit, eine Reputation aufzubauen ausgeschlossen, da sie einmalig und unter der Bedingung vollständiger Anonymität stattfinden, ganz abgesehen von der meist fehlenden Möglichkeit, die Reputation glaubwürdig zu signalisieren.

#### **3.2.4. Strong Reciprocity, Gruppenselektion, Mehrebenen-selektion und Ko-Evolution von Genen und Kultur**

Gintis (2000) beschreibt ein evolutionäres Modell, das die Entstehung altruistischer Bestrafung erklären kann. Neben reziprok altruistischen Individuen (s. o.) existieren in diesem Modell auch stark reziproke Individuen. Sie unterscheiden sich von den reziprok altruistischen Individuen dadurch, dass sie nicht nur kooperatives Verhalten mit kooperativem beantworten und unkooperatives mit unkooperativem, sondern sie bestrafen obendrein unkooperative Individuen und belohnen kooperative. Auf diese Weise wird berücksichtigt, dass in Situationen, die das Fortbestehen einer Gemeinschaft bedrohen, reziprok altruistisches Verhalten nicht mehr ausreicht, um die Kooperation aufrecht zu erhalten. Die reziprok altruistischen Individuen können nicht davon ausgehen, dass eine altruistische Handlung heute sich morgen noch auszahlt, daher werden diese in einer solchen Situation unterlassen. Ist aber nach wie vor eine

genügend große Zahl an stark reziproken Individuen bereit, Normverletzungen zu bestrafen, kann auch in solchen Situationen die überlebenswichtige Kooperation aufrecht erhalten werden und die ausschließlich eigennutzinteressierten Mitglieder der Gruppe können zur Kooperation bewegt werden. Hier ergibt sich allerdings das gleiche Problem, das weiter oben bereits bei den sozialen Normen angesprochen wurde: Individuen, die zwar kooperieren, aber nicht belohnen oder bestrafen, haben einen Fitnessvorteil gegenüber stark reziproken Individuen. Ein öffentliches Gut zweiter Ordnung – die Bestrafung<sup>8</sup> – entsteht (Yamagishi, 1986) und somit auch die Möglichkeit des Trittbrettfahrens zweiter Ordnung. (Kooperation in Bezug auf das Öffentliche Gut erster Ordnung, Nicht-Kooperation in Bezug auf die Bestrafung / Belohnung derer, die nicht zum eigentlichen öffentlichen Gut erster Ordnung beigetragen haben) (vgl. Heckathorn, 1989).

Eine Möglichkeit, das Rätsel individueller Fitnessnachteile durch die Bestrafung aufzulösen, besteht darin, Selektionsprozesse (auch) auf der Ebene der Gruppe anzusiedeln (Boyd, Gintis, Bowles & Richerson, 2003; Gintis, Bowles, Boyd & Fehr, 2003; Sober & Wilson, 1998). Dieser Ansatz ist allerdings umstritten. Die Argumentationskette ist wie folgt aufgebaut: Aufgrund eines Anteils von altruistisch bestrafenden (belohnenden) Individuen (starke Reziprozität) kommt es in einer Gruppe zu einem hohen Maß an Kooperation (z.B. Beitragen zu Gemeinschaftsgütern). Diese Kooperationsgewinne stellen einen Vorteil gegenüber Gruppen dar, in denen nicht kooperiert wird. Gruppen mit stark reziproken Akteuren hätten dementsprechend einen Selektionsvorteil auf Gruppenebene (Boyd et al., 2003).

Das Problem des negativen Selektionsdrucks aufgrund der individuellen Fitnessnachteile durch Kooperation und Bestrafung wird im Modell so gelöst: Generell gilt, dass die Individuen, die sich nicht an der Bereitstellung der Gemeinschaftsgüter beteiligen, einen Vorteil

---

<sup>8</sup> Auch Schelling (1978) verweist bereits darauf, dass Bestrafung ein öffentliches Gut sein kann.

gegenüber den kooperativen (reziproken) Individuen haben und diese haben wiederum einen Vorteil gegenüber den bestrafenden und belohnenden (stark reziproken) Individuen. Der Vorteil der Trittbrettfahrer wird durch die Bestrafung verringert, was aber passiert mit den nicht-bestrafenden Kooperatoren? Ihr relativer Vorteil wird ebenfalls geringer, weil die Kosten und somit die Fitness-Nachteile der Bestrafenden abnehmen, wenn die Rate der Kooperation und die Anzahl der Bestrafenden zunehmen (Gürerk, Irlenbusch & Rockenbach, 2006; Boyd et al., 2003; Gintis et al., 2003). Im Gegensatz zur Belohnung ungerechten Verhaltens, deren Kosten mit dem Grad der Kooperation steigen, gehen die Kosten der Bestrafung mit dem Grad der Kooperation zurück. Kooperieren alle, haben die stark reziproken Individuen keinen Nachteil mehr.

Modelle der Gene-Culture-Coevolution akzeptieren grundsätzlich den Gedanken der Selektion auf der Ebene des Individuums, gehen aber zusätzlich davon aus, dass auf der Ebene der Gruppe kulturelle evolutionäre Prozesse stattfinden, die den Selektionsdruck auf der Individualebene schwächen (Henrich & Boyd, 2001). Dies können z.B. Lernprozesse sein, bei denen das Verhalten besonders erfolgreicher Gruppenmitglieder oder das der Mehrheit kopiert wird (Modell-Lernen).

### **3.2.5. Das Commitment-Modell**

Ein Modell, das sowohl die Entstehung moralischer Emotionen erklären kann als auch das Problem individueller Fitness-Nachteile durch die Bestrafung auflöst, ist das Commitment-Modell (Frank, 1988). Hier wird argumentiert, dass Menschen, die sich fair und altruistisch verhalten, von anderen erkannt und aktiv als Interaktions- und Sexualpartner ausgesucht werden, so dass ihnen ein Vorteil entsteht. Menschen sind nach dem Modell von Frank besonders in wichtigen Situationen gut darin, Betrüger zu erkennen. Wenn es schwer, kostspielig und riskant ist, unwahre Signale über die eigene Ehrlichkeit zu senden, lohnt es sich immer, ein guter Mensch zu sein. Also sind

die fairen Akteure im Modell von Frank auch dann fair, wenn niemand zuschaut, denn die einfachste Möglichkeit, gut zu erscheinen, ist, gut zu sein. Frank betont zudem die entscheidende Rolle moralischer Emotionen, die sich im Rahmen der Evolution gebildet haben, weil sie helfen, das Commitment-Problem menschlicher Interaktion zu lösen. Frank sieht Gefühle und Emotionen als proximate Auslöser von Verhalten, und bestimmte moralische Emotionen bringen Kooperationsrateure dazu sich kooperativ zu verhalten. Diese Signale können von potentiellen Interaktionspartnern erkannt werden, so dass sie häufiger für (gewinnbringende, kooperative) Interaktionen ausgewählt werden. Allerdings ist das Erkennen der Kooperationsbereitschaft bzw. die Suche nach kooperativen Individuen mit Kosten verbunden.

Wenn es in einer Gruppe viele unkooperative Individuen gibt, lohnt es sich, die Kosten auf sich zu nehmen, kooperative Individuen aufzuspüren und mit ihnen den erhöhten Nutzen kooperativer Interaktion zu teilen. Aufgrund dieses erhöhten Ertrags, den die Kooperativen erwirtschaften, können sie sich stärker vermehren, denn der Ertrag, den zwei unkooperative Individuen erwirtschaften, ist deutlich niedriger (s. u. die Beispiele zum Gefangenenden-Dilemma). Wenn der Anteil kooperativer Individuen ein bestimmtes Niveau erreicht hat, haben unkooperative Individuen wieder einen Vorteil, da sie eine gute Chance haben, mit ihrer unkooperativen Strategie auf ein kooperatives Individuum zu treffen und so ihren eigenen Ertrag zu maximieren. Mit Hilfe des Commitment-Modells lässt sich also auch erklären, warum es in einer Gesellschaft sowohl kooperative als auch unkooperative Individuen gibt und warum die Evolution nicht zwangsläufig zum Aussterben einer der beiden Gruppen führen muss.

Damit Menschen die richtigen Interaktionspartner entsprechend deren Prosozialität auswählen können, müssen die altruistischen Merkmale einer Person valide erkannt werden können (Fetchnhauer & Bierhoff, 2004), was wiederum voraussetzt, dass Prosozialität eine



verhältnismäßig stabile Persönlichkeitsdisposition ist (Hinweise darauf gibt Bierhoff, 2002). In die gleiche Richtung weisen auch verschiedene Studien zum Einfluss von Ungerechtigkeitssensibilität auf pro- und antisoziales Verhalten und die Verbindung zwischen Ungerechtigkeitssensibilität und anderen Persönlichkeitsmaßen (Fetchenhauer & Huang, 2004; Gollwitzer et al., 2005; Schmitt et al., 2005). Hinweise darauf, dass Menschen in der Lage sind, kooperative Mitmenschen mit überzufälliger Wahrscheinlichkeit zu identifizieren, liefern die Untersuchungen von Frank, Gilovich und Regan (1993) sowie Fetchenhauer und Groothuis (2007).

### **3.2.6. Alte Verhaltensweisen (fälschlich) angewandt auf neue Situationen**

Es wird verschiedentlich darauf hingewiesen, dass sich die Umgebung der Menschen im Verlauf der Evolution stark verändert hat (bezeichnet als EEA: „Environment of Evolutionary Adaptivness“). Früher lebten Menschen eher in kleinen Gruppen (Sippe, Familie) zusammen. Daher könnte es sein, dass das aus Sicht der individuellen bzw. Verwandtschafts-Selektion unadaptive Verhalten in einmaligen Dilemmasituationen, nämlich altruistisch zu bestrafen oder sich altruistisch zu verhalten, daher rührt, dass bei einmaliger Interaktion die gleichen Mechanismen (insb. moralische Emotionen) ausgelöst werden wie bei wiederholter Interaktion (Camerer & Fehr, 2004, S. 78 berichten, dass andere diese "unnatural habit theory" hätten, nennen aber keine Quellen; Hagen & Hammerstein, 2006, S. 342 diskutieren verschiedene Formen einer ähnlichen "mismatch hypothesis"). Selbst wenn Menschen zwischen den Situationen unterscheiden können, wofür es überzeugende Hinweise gibt (Fehr & Fischbacher, 2003), würden sie dann in beiden Situationen ähnlich handeln, z.B. weil die gleichen Emotionen ausgelöst werden (Hagen & Hammerstein, 2006). Im Fall einmaliger Interaktion wäre dies auf individueller Ebene maladaptiv, hätte jedoch im Verlauf der Evolution nicht zur Extinktion geführt, da solche Interaktionen in der Menschheitsgeschichte ver-

mutlich eher selten vorkamen (Gintis et al., 2003, widersprechen dieser Ansicht).

### **3.2.7. Zusammenfassung der ultimatzen Erklärungen**

Die Diskussion der ultimatzen Gründe für altruistisches Verhalten wird immer noch offen geführt. Verschiedene Theorien sind in der Lage, Teile der Entstehung altruistischen Verhaltens (z.B. gegenüber Verwandten oder bei sich wiederholenden Interaktionen) zu erklären. Menschen kooperieren jedoch offensichtlich auch mit nicht-verwandten Individuen (in der Realität und in Experimenten), daher führen Theorien, die rein auf Verwandtschaftsaltruismus basieren, nicht weiter. Wenn es darum geht, Verhalten in einmaligen Situationen zu erklären, scheiden auch die Theorien aus, die eine Fortsetzung der Interaktion annehmen. Deshalb kommen sowohl reziproker Altruismus als auch indirekter Altruismus sowie Reputation als Erklärung für das Verhalten in einmaligen anonymen Interaktionen nicht in Frage (vor allem nicht bei der Bestrafung durch unbeteiligte Dritte).

Eine gute Erklärungsgrundlage für altruistisches Verhalten bietet das Commitment-Modell von Frank (1988). Es kann zum einen erklären, warum in einer Gesellschaft gleichzeitig kooperative und unkooperative Individuen zu finden sind, als auch warum keine der beiden Gruppen im Laufe der Evolution ausgestorben ist. Außerdem ist die Erklärung der Vorteile kooperativer Individuen bei der Interaktion mit anderen kooperativen Individuen gegenüber der Interaktion von unkooperativen Individuen schlüssig. Verschiedentlich konnte auch gezeigt werden, dass Individuen überzufällig gut erkennen können, ob ein anderes Individuum kooperativ ist oder nicht. Damit kann erklärt werden, dass kooperative Individuen bei geeigneter Wahl der Interaktionspartner keine Fitnessnachteile erleiden müssen. Darüber hinaus schlägt das Modell eine tragfähige Brücke zur proximatzen Erklärung des Verhaltens durch moralische Emotionen. Die moralischen Emotionen sind der Auslöser für kooperatives Verhalten bzw.

für die Bestrafung unkooperativen Verhaltens. Ihr Entstehen kann mit deren Nützlichkeit als Signal der Kooperationsbereitschaft erklärt werden, was seinem Träger häufige lohnende Interaktionen mit anderen kooperativen Individuen ermöglicht.

Im nun folgenden empirischen Teil dieser Arbeit werden Teile dieser Argumentation empirisch überprüft, indem zum einen der Einfluss von Persönlichkeitseigenschaften und moralischen Emotionen auf das Bestrafungsverhalten untersucht und zum anderen die Frage geklärt wird, ob Menschen die Bestrafung durch andere antizipieren können und entsprechend darauf reagieren.

## **4. Altruistische Bestrafung in experimentellen Spielen**

Ein bestimmtes Verhalten in der realen Welt zu beobachten, ist eine Sache. Dieses Verhalten unter kontrollierten Bedingungen zu messen, um Aussagen über Ursachen und Wirkungen zu treffen, eine andere. Mit der einfachen und abstrakten Darstellung von sozialen Interaktionen in „Spielen“ liefert die experimentelle Spieltheorie ein gut handhabbares Instrument zur Messung altruistischer Bestrafung.

### **4.1. Spieltheorie und Spiele**

#### **4.1.1. Spieltheorie als Sprache und Instrument experimenteller Ökonomen**

Das Erscheinen des Buches „The Theory of Games and Economic Behavior“ von John von Neumann und Oskar Morgenstern (1944) wird allgemein als der Beginn der Disziplin „Spieltheorie“ gesehen. In der Spieltheorie wird versucht, das Handeln verschiedener Spieler, die in bestimmten Situationen (den Spielen) nach bestimmten Regeln interagieren, mathematisch abzuleiten. Das Werkzeug, das zur Analyse und Vorhersage genutzt wird, sind die Spiele, die im Prinzip nur die Beschreibung einer Situation sind (Spieler, Strategien, Regeln). Spieltheorie kann in diesem Sinn als Sprache verstanden werden, mit der bestimmte Situationen beschrieben werden können. Diese universelle Sprache, die am Anfang nur Ökonomen und Mathematiker gesprochen und verstanden hatten, verbreitete sich in den vergangenen Jahren auch in anderen Disziplinen wie Psychologie, Soziologie und Biologie. Die Untersuchung dessen, was echte Spieler tatsächlich tun, und die Berücksichtigung von Emotionen, Fehlern, Lernen usw., wird auch als „*behavioral game theory*“ bezeichnet (Camerer, 2003).

Auf der einen Seite besteht die Spieltheorie aus der mathematischen Ableitung exakter Gleichgewichte und theoretischer Vorhersagen des

Verhaltens von Spielern auf der Basis der Annahme, dass sie eigeninteressiert nur ihren Nutzen maximieren (gemessen ausschließlich in Geldeinheiten, die sie im Spiel verdienen). Auf der anderen Seite können die Spiele der Spieltheorie aber auch genutzt werden, um menschliches Verhalten in klar definierten Situationen, ohne die mathematische Ableitung von Gleichgewichten, experimentell zu untersuchen.<sup>9</sup> Im Allgemeinen sind experimentelle Ökonomen an einer ganz speziellen Form der Interaktion interessiert. Sie untersuchen meist die einmalige Interaktion zwischen Fremden, die sich nicht kennen und auch nicht kennen lernen werden. Dabei geht es fast immer um echtes Geld und zwischen den Spielern gibt es keine Kommunikation. Diese Basissituation dient als Referenzmarke, wenn z.B. der Einfluss wiederholter Interaktionen untersucht werden soll. Die Spielsituation wird in normaler, nicht-wertender Sprache beschrieben (Worte wie „fair“, „bestrafen“ etc. werden vermieden), um die Versuchspersonen nicht zu beeinflussen (Camerer & Fehr, 2004).

In dieser Arbeit geht es nicht um die mathematische Herleitung spieltheoretischer Gleichgewichte. Es geht auch nicht um die Modellierung oder Simulation des Verhaltens künstlicher Probanden. Hier wird das Instrumentarium der Spieltheorie – die Spiele – als Messinstrument für altruistische Bestrafung eingesetzt. Es wird bei den einzelnen Spielen erläutert, zu welchen Voraussagen die Spieltheorie kommt, um einen Vergleichspunkt zu haben. Ansonsten werden die Spiele aber lediglich als Werkzeuge betrachtet. Um zu zeigen, dass sie hervorragend geeignet sind, altruistische Bestrafung zu messen, werden bei der Vorstellung der einzelnen Spiele die empirischen Befunde kurz dargestellt.

Da die Zahl der Spiele mittlerweile sehr hoch ist, sei für eine erschöpfende Darstellung verschiedener Spiele an Camerer (2003) verwiesen. Hier soll lediglich auf die Spiele eingegangen werden, die direkt zur Messung von altruistischer Bestrafung verwendet werden,

---

<sup>9</sup> Auf die Eignung einfacher Spiele zur Messung von sozialen Motiven wurde bereits recht früh verwiesen (Marwell & Schmitt, 1968).

und auf solche, die zur Erklärung nötig sind. (Für einen allgemeinen Überblick über Spiele zur Messung sozialer Normen und Präferenzen siehe Camerer & Fehr, 2004)<sup>10</sup>.

Nach einer allgemeinen Charakterisierung experimenteller Spiele werden dann u. a. das Ultimatum-Spiel, das Diktator-Spiel, das Drei-Personen-Bestrafungs-Spiel, das Gefangenen-Dilemma, das Vertrauensspiel und das Öffentliche-Gut-Spiel vorgestellt. Der Schwerpunkt liegt auf den ersten drei Spielen, die auch in den eigenen empirischen Untersuchungen verwendet wurden. Zunächst sollen nun einige allgemeine Charakteristika experimenteller Spiele erläutert werden.

#### **4.1.2. Grundlegende Charakteristika experimenteller Spiele**

##### *Anzahl der Spieler*

Verschiedene Spiele sehen eine unterschiedliche Anzahl von Spielern vor. Aufgrund der Implikationen der verschiedenen Spielerzahlen, ist es sinnvoll zwischen Zwei-, Drei- und Mehr-Personen-Spielen zu unterscheiden. Grundsätzlich sind der Anzahl der Spieler keine Grenzen gesetzt (vgl. weiter unten Öffentliches-Gut-Spiel). Die im experimentellen Teil der Arbeit untersuchten Spiele sind ausschließlich Zwei- und Drei-Personen-Spiele, deshalb wird bei der Vorstellung der Spiele hier auch der Schwerpunkt gesetzt.

##### *Wiederholt vs. Einmalig*

Ein weiterer Unterscheidungspunkt ist die Zahl der Wiederholungen, der einzelnen Runden, der Spiele. Hier kann man zwischen einmaligen (one-shot) und wiederholten (repeated) Spielen unterscheiden.

Bei wiederholten Spielen kann entweder mehrfach mit dem gleichen Partner interagiert werden (Partner-Bedingung) oder man bekommt in jeder Runde zufällig einen neuen Partner zugewiesen (Stranger-Bedingung) (Camerer & Fehr, 2004).

---

<sup>10</sup> Der Schwerpunkt liegt in der vorliegenden Arbeit auf Bestrafung. Natürlich gibt es auch Spiele, die eine Belohnung nach fairen oder großzügigen Handlungen modellieren (Fehr & Fischbacher, 2002).

Spielt man verschiedene Runden mit den gleichen Partnern, sind strategische Interaktionen denkbar. Z.B. kann ein Spieler sich eine Reputation als kooperativer Spieler oder als jemand, der Normabweichungen streng bestraft, aufbauen. In diesem Fall könnte man von einer Bestrafung der Mitspieler in den Folgeperioden (materiell) profitieren. Spielt man hingegen immer wieder mit neuen Partnern, ist eine Bestrafung aus strategischer eigennütziger Sicht nicht sinnvoll, denn die Bestrafung nützt dem Bestrafenden nicht. Selbst wenn sie erfolgreich sein sollte, kann der Bestrafende nicht von der zukünftigen Normkonformität des Bestraften profitieren, weil er in weiteren Runden einen anderen Partner haben wird. Andere profitieren aber schon, so dass die Bestrafung unter diesen Umständen als altruistischer Akt gesehen werden kann. Fehr und Fischbacher (2002) berichten von einem Vergleich zwischen Partner- und Stranger-Bedingung, der keine Beweise für strategisches Bestrafen liefert. In beiden Bedingungen sind die Strafen nahezu gleich hoch. Die Wiederholung der Spiele führt also zu (un-)erwünschten Nebeneffekten. In wiederholten Spielen lassen sich Lerneffekte der Probanden sowie strategische Gesichtspunkte untersuchen (für eine Übersicht siehe Camerer, 2003; Roth, 1995).

### *Anonymität*

Die meisten Experimente in der experimentellen Ökonomie werden unter Anonymität durchgeführt. Dies bedeutet in der Regel, dass ein Spieler nicht weiß, wer sein direkter Interaktionspartner in einem Spiel ist. Anonymität zwischen den Versuchspersonen ist wichtig, da es einen Einfluss auf das Spielverhalten allgemein und auf die Art und Höhe der altruistischen Bestrafung haben kann, wenn die Versuchspersonen sich kennen. Auf Anonymität wird im Allgemeinen nur verzichtet, wenn explizit das Verhalten von Personen untersucht werden soll, die sich kennen bzw. miteinander kommunizieren (Camerer, 2003).

Bei der Anonymität sind verschiedene Grade denkbar:

1. Die Versuchspersonen kennen sich nicht, haben sich noch nie gesehen und werden sich auch nicht kennen lernen/sehen.
2. Die Versuchspersonen kennen sich nicht, haben sich aber vor dem Experiment kurz gesehen und sitzen mit anderen Spielern in einem gemeinsamen Computerlabor.
3. Die Versuchspersonen kennen sich.

Meistens werden die Experimente in größeren Computerlaboren durchgeführt (ca. 30 Versuchspersonen in einem Labor). Die Versuchspersonen kennen sich zwar meist nicht persönlich, sitzen aber mit ihren Mitspielern in einem Raum.

Eine weitere Verschärfung der Anonymität besteht darin, den Versuch so durchzuführen, dass der Versuchsleiter nicht weiß bzw. nicht wissen kann, wie sich eine einzelne Versuchsperson verhalten hat. Es herrscht also Versuchsperson-Versuchsleiter-Anonymität (double blindness) (vgl. z.B. Bolton & Zwick, 1995). Während die Frage der Anonymität zwischen den Versuchspersonen große Auswirkungen auf die Ergebnisse hat, bleiben die Ergebnisse von der Versuchsperson-Versuchsleiter-Anonymität weitgehend unbeeinflusst (Roth, 1995).

#### *(Monetäre) Anreize*

Für gewöhnlich werden in experimentellen Spielen reale monetäre Anreize gegeben, d.h., die Versuchspersonen spielen meist um Geld bzw. müssen Geldbeträge unter sich aufteilen. Für die meisten Menschen kann nämlich angenommen werden, dass – zumindest in den Bereichen in denen sich die Experimente bewegen – keine „Sättigung“ eintritt, so dass die Versuchspersonen einen größeren Geldbetrag immer mehr schätzen als einen geringeren (Camerer, 2003). Hertwig und Ortmann (2001) nennen darüber hinaus die hohe Salienz monetärer Anreize und die Tatsache, dass in Experimenten der experimentellen Ökonomie meist die ökonomische Theorie getestet werden soll, wofür sich natürlich monetäre Anreize anbieten.



Einen Überblick über die Implikationen monetärer Anreize geben Camerer und Hogarth (1999). Die Befundlage ist allerdings gemischt (siehe auch Hertwig & Ortmann, 2001). Allgemein scheint es aber so zu sein, dass monetäre Anreize zu einer Verbesserung der Leistung (geringere Variabilität der Daten, weniger Ausreißer) in Experimenten führen (Hertwig & Ortmann, 2001). Auf das tatsächliche Verhalten (z.B. Annahme / Ablehnung im Ultimatum-Spiel) scheint die Höhe der Bezahlung allerdings keinen großen Einfluss zu haben (Camerer, 2003). Über die negativen Folgen zu geringer – beleidigender – Bezahlung berichten Gneezy und Rustichini (2000) in ihrem Artikel „Pay Enough or don't pay at all“.

Häufig werden die Spiele nicht direkt mit Geld gespielt, sondern mit Punkten, die nach dem Spiel in Geld umgerechnet werden. Ein Vorteil dieser Vorgehensweise ist, dass bei interkulturellen Vergleichen die Punktzahlen gleich bleiben können und dass über einen entsprechenden Umrechnungsfaktor im Nachhinein sichergestellt werden kann, dass die Versuchspersonen in den verschiedenen Ländern um vergleichbare Summen gespielt haben. Ein Nachteil dieser Vorgehensweise könnte darin gesehen werden, dass das Spielen um Punkte recht abstrakt ist und die daraus folgenden Anreize unter Umständen verzerrt werden. Ein systematischer Vergleich von Bezahlung in Punkten mit der Bezahlung in bar wurde m. W. bisher noch nicht vorgenommen. Insbesondere müsste bei einem solchen Versuch bei der Barzahlung auch noch zwischen einer Bezahlung vor dem Versuch und nach dem Versuch unterschieden werden.

*„No Deception“ – keine falschen Angaben gegenüber den Versuchspersonen*

In der (Sozial-)Psychologie ist es durchaus üblich, die Versuchspersonen (zunächst) über den wahren Grund oder bestimmte Aspekte der Untersuchung im Unklaren zu lassen bzw. explizit zu täuschen<sup>11</sup>. Ein solches Verfahren widerspricht dem No-Deception-Grundsatz der

---

<sup>11</sup> Später erfolgt dann eine Aufklärung, die insbesondere dazu dient, die Versuchspersonen ohne psychische Beeinträchtigung aus dem Experiment zu entlassen.

experimentellen Ökonomie, der besagt, dass die Versuchspersonen stets voll über alle Aspekte des Versuchs informiert sein müssen. Dies hat weniger moralische Gründe, sondern soll vielmehr dazu dienen, dass die Versuchspersonen bei weiteren Experimenten nicht argwöhnisch sind. Gelegentlich wird jedoch von der No-Deception-Regel abgewichen, wenn gute Gründe für ein solches Verhalten vorliegen. In der Vergangenheit lassen sich verschiedene Studien mit spieltheoretischem Design finden, bei denen Versuchspersonen gegen simulierte oder eingeweihte Gegner spielten, wobei die Instruktionen vorgaben, die Gegner seien real bzw. echte Versuchspersonen (vgl. z.B. Diekmann, 2004; Horne, 2004; Kravitz & Gunto, 1992; Roth & Schoumaker, 1983, S. 367: "The instructions led them to believe they were bargaining with other individuals"; Harrison & McCabe, 1996, geben an, die gleiche Verfahrensweise zu verwenden wie Roth & Schoumaker; Singer et al., 2006, nutzen professionelle Schauspieler als eingeweihte Spieler, sog. „Confederates“). Auf diese Weise lassen sich die Reaktionen der realen Spieler auf bestimmte Angebote in experimentellen Spielen kontrolliert untersuchen. In Studien, die aufwändige bildgebende Verfahren der Hirnforschung nutzen, wird teilweise unter Verweis auf die großen logistischen Probleme ein ähnliches Verfahren angewendet (vgl. Sanfey et al., 2003). Ein weiterer Grund für die Simulation der anderen Interaktionspartner kann darin bestehen, dass ein Verhalten, das nur selten vorkommt, dennoch unter realistischen Bedingungen gemessen werden soll.

In anderen Fällen werden die Versuchspersonen über weitere Runden bzw. Experimente im Unklaren gelassen, um die Ergebnisse verschiedener Runden eines Spiels nicht zu vermischen (vgl. Fehr & Fischbacher, 2004b). Darüber, ob eine nicht gegebene Information auch eine Form von Täuschung ist, gibt es unterschiedliche Meinungen. Hertwig und Ortmann (2001) sehen darin keine Täuschung, in den dem Artikel folgenden Kommentaren werden hierzu jedoch auch gegenteilige Meinungen geäußert. Einen Überblick über die Nachteile von Täuschungen in Experimenten geben Hertwig und Ortmann

(2001; im Anschluss an den Artikel finden sich verschiedene Kommentare, in denen die Vor- und Nachteile diskutiert werden).

Die vorangegangenen Beispiele verdeutlichen, dass es durchaus Fälle gibt, in denen die Versuchspersonen über Teile des Experiments nicht vollständig oder sogar falsch informiert wurden. Dies geschieht aus verschiedenen methodischen, finanziellen oder logistischen Gründen. Abschließend bleibt festzuhalten, dass die No-Deception-Regel nicht leichtfertig aufgegeben werden und dass man eine Abwägung zwischen Kosten und Nutzen des Vorgehens vornehmen sollte. Die Abwägung wird auch von der finanziellen Ausstattung des entsprechenden Forschers abhängen. Die No-Deception-Regel sollte nicht dazu führen, dass finanziell weniger gut ausgestattete Forscher bestimmte Untersuchungen nicht mehr durchführen können. In der vorliegenden Arbeit wird z.B. der Vorgehensweise von Roth & Schoumaker (1983) gefolgt, da der finanzielle und logistische Aufwand der Untersuchungen sonst unverhältnismäßig hoch gewesen wäre.

#### **4.2. Second-Party-Punishment (SPP)**

Unter Second-Party-Punishment (im Weiteren abgekürzt mit SPP) versteht man die Bestrafung des Normverletzers durch eine beteiligte, selbst betroffene Person. Das bedeutet, dass das unkooperative, unfaire Verhalten eines Spielers einen direkten negativen Einfluss auf den Ertrag des bzw. der anderen Spieler hat (Camerer & Fehr, 2004), die ihn bestrafen (können). Hiervon abzugrenzen ist das Third-Party-Punishment (TPP), das weiter unten behandelt wird. Die Trennung erfolgt, weil es Unterschiede in den Ursachen (z.B. andere auslösende Emotionen: Ärger vs. Empörung) und der Höhe (SPP ist meist stärker als TPP) der beiden Bestrafungsarten gibt. Im nächsten Abschnitt werden Spiele und empirische Ergebnisse zum SPP vorgestellt.

#### **4.2.1. Ultimatum-Spiel (Ultimatum Game)**

Stellen wir uns einen Verkäufer in einem kleinen Gemüseladen in Familienbesitz vor. Dieser Verkäufer bietet samstags Erdbeeren an, von denen er weiß, dass er sie am Montag nicht mehr verkaufen kann. Kurz vor Ladenschluss kommt ein Tourist und bietet an, die Erdbeeren, die einen Warenwert von 10 Euro haben, für einen Euro zu kaufen. Eigentlich müsste der Verkäufer dem Angebot zustimmen, denn ein Euro ist mehr als kein Euro, und wenn er es ablehnt, haben beide nichts. Ob er aber tatsächlich zustimmt, ist eine andere Frage, denn der Tourist hat von diesem Geschäft deutlich mehr, und das könnte der Händler als ungerecht auffassen.

Eine solche Situation, in der es um die Aufteilung einer Ressource zwischen zwei Personen geht, kann mit dem Ultimatum-Spiel (Güth et al., 1982) modelliert werden. Am Ultimatum-Spiel nehmen zwei Spieler teil. Einer der Spieler – der Aufteilende – erhält einen Geldbetrag  $S$  vom Versuchsleiter, den er zwischen sich und dem zweiten Spieler – dem Entscheider – aufteilen soll. Für gewöhnlich wird den Spielern zufällig eine von zwei Rollen zugewiesen. Manchmal – wie in einigen der weiter unten geschilderten Studien – kann es sinnvoll sein, die Spieler gezielt den Rollen zuzuordnen. Der Aufteilende macht dem Entscheider ein Angebot in Höhe von  $x$ . Der Entscheider hat die Wahl, das Angebot anzunehmen oder abzulehnen. Wenn der Entscheider das Angebot  $x$  annimmt, erhält er diesen Betrag  $x$ , und der Aufteilende erhält den anfänglichen Geldbetrag  $S$  minus  $x$ . Lehnt der Entscheider das Angebot ab, erhalten beide nichts. Der Entscheider setzt also in diesem Fall eigene Ressourcen ein, um einen unfairen Partner zu bestrafen. Mit Hilfe des Ultimatum-Spiels kann daher die Bereitschaft gemessen werden, einen unfairen Partner unter Einsatz eigener Ressourcen zu bestrafen. Bei Ablehnung des Aufteilungsvorschlags werden die persönlichen Kosten sehr deutlich. Lehnt ein Entscheider bei der Aufteilung von 10 Euro eine 8:2-Aufteilung ab, kostet ihn dies also 2 Euro (Camerer & Fehr, 2004). Es wird auch

ein weiterer großer Vorteil der Messung durch experimentelle Spiele klar: Die Kosten sind direkt monetär messbar.

Weitere Zahlenbeispiele:

1. Der Aufteilende schlägt eine Aufteilung vor, die vorsieht, dass er von insgesamt 10 Euro 5 Euro behält und 5 Euro an den Entscheider abgibt. Der Entscheider akzeptiert und beide erhalten entsprechend der Aufteilung 5 Euro.
2. Der Aufteilende schlägt eine Aufteilung vor, die vorsieht, dass er die gesamten 10 Euro behält und 0 Euro an den Entscheider abgibt. Der Entscheider lehnt dies ab und beide erhalten 0 Euro.

Die Spieltheorie sagt voraus, dass ein rationaler, eigennutzorientierter Entscheider jeden Betrag annehmen würde, der größer als Null ist. Außerdem zieht jeder Spieler einen größeren Geldbetrag einem kleineren vor. Der Aufteilende weiß das und bietet daher den kleinstmöglichen Betrag an (z.B. einen Cent). Das Angebot würde nicht abgelehnt werden, da ein Cent für den Entscheider besser als Null Cent ist (Camerer & Fehr, 2004).

Wie lässt sich das Ultimatum-Spiel psychologisch interpretieren? Da die Spieler meist den unterschiedlichen Rollen zugewiesen werden, gibt es eigentlich keinen Grund, warum ein Spieler mehr erhalten sollte als ein anderer. Verschiedene distributive Gerechtigkeitsprinzipien weisen zudem in die gleiche Richtung: Bei der Verteilung eines Geldbetrages unter zwei Personen, die den gleichen Einsatz erbracht haben (im Fall der Losentscheidung haben sie beide keinen Einsatz gebracht), sollte man sowohl nach dem Equity-Prinzip als auch nach dem Gleichheitsprinzip den Betrag gleich aufteilen. Das Bedürfnisprinzip wäre nur dann sinnvoll anwendbar, wenn die Bedürftigkeit der einzelnen Spieler festgestellt werden könnte. Es gibt aus gerechtigkeitspsychologischer Sicht keinen Grund für eine Aufteilung, die von der Gleichverteilung abweicht (vgl. z.B. Walster et al., 1978; in gleicher Weise argumentieren auch Fehr & Schmidt, 1999, explizit

in Bezug auf experimentelle Spiele; Bolton & Ockenfels, 2000, sprechen von der 50/50-Verteilung als sozialem Referenzpunkt; Fetschenhauer & Huang, 2004). Ob ein Aufteiler nun fair oder unfair aufteilt, wird u. a. von der Stärke seiner moralischen Überzeugungen bzw. seiner Persönlichkeit abhängen. Die Frage ist, ob er eine Norm internalisiert hat, sich zu jeder Zeit fair und gerecht zu verhalten oder ob es bei ihm in bestimmten Bereichen zu einer Abwägung zwischen Gerechtigkeit und Eigennutz kommt, z.B. wenn wie hier die negativen Folgen ungerechten Handelns nicht sonderlich gravierend sind.

Was ist die psychologische Vorhersage für den Entscheider? Wenn die 50/50-Verteilung als einzige gerechte Verteilung wahrgenommen wird, führt eine Abweichung nach der Equity-Theorie zu einem negativen Spannungszustand bzw. Ärger (Hassebrauck, 1984) beim Entscheider, der auf Beseitigung drängt. Ob Ärger empfunden wird und wie stark dieser Ärger ist, wird unter anderem davon abhängen, wie stark man auf eine ungerechte Aufteilung reagiert (vgl. die Ausführungen zur Ungerechtigkeitssensibilität weiter oben). Eine Beseitigung des Ärgers kann (zumindest für einige Personen) durch das Ablehnen eines unfairen Angebotes erreicht werden. Je ungerechtigkeitssensibler ein Empfänger ist, desto mehr sollte ihm die Wiederherstellung der Gerechtigkeit wert sein (Fetschenhauer & Huang, 2004).

Ein Aufteiler, der sich im Ultimatum-Spiel neun von zehn Euro nimmt und dem Empfänger nur einen Euro zubilligt, fügt dem Empfänger nicht nur einen materiellen Schaden zu (gemessen an der Gleichverteilungsnorm). Zugleich drückt er gegenüber dem Entscheider (und der Gesellschaft) aus, dass er sich offensichtlich nicht an diese Norm gebunden fühlt, dass er über dem Entscheider steht und anderes. Der Schaden für das Opfer und die Gesellschaft ist also größer als der materielle Schaden. Aus diesem Grund sind Modelle, die sich ausschließlich auf den materiellen Schaden beziehen, nur schlecht in der Lage, alle Reaktionen auf einen solchen

Normbruch zu erklären. Eine Bestrafung dient dann (auch) dazu, das psychologische Gleichgewicht wiederherzustellen, indem sie das Selbstbild und die Werte des Bestrafenden wiederherstellen (Miller & Vidmar, 1981).

Bereits die ersten Experimente zum Ultimatumspiel zeigten, dass sich die Versuchspersonen nicht entsprechend den spieltheoretischen Voraussagen verhielten. „[...] subjects often rely on what they consider a fair or just justified result. Furthermore, the ultimatum aspect cannot be completely exploited since subjects do not hesitate to punish if their opponent asks for 'too much'” (Güth et al., 1982, S. 384). Insgesamt ist die Befundlage für die spieltheoretische Voraussage negativ (vgl. z.B. Bolton & Zwick, 1995). Henrich et al. (2005, S. 797) fassen ihre Ergebnisse aus 15 Stammesgesellschaften aus der ganzen Welt bezüglich des Eigennutzaxioms der ökonomischen Theorie mit folgendem Satz zusammen: „[...] there is no society in which experimental behavior is fully consistent with the selfishness axiom“.

Viele Untersuchungen haben gezeigt, dass der Aufteilende im Durchschnitt ca. 40 bis 50 Prozent des aufzuteilenden Betrages an den Entscheider gibt. Aufteilungen von bis 10 Prozent kommen so gut wie nicht vor, genauso wenig wie überfaire Angebote von 51 bis 100 Prozent. Sehr niedrige Angebote bis ca. 20 Prozent werden in der Regel in der Hälfte der Fälle abgelehnt (Camerer, 2003). Ein Vergleich mit dem Diktator-Spiel im nächsten Abschnitt zeigt allerdings, dass viele faire Angebote eher durch die Angst vor Ablehnung zustande kommen und weniger durch intrinsische Fairness der Aufteiler.

Verschiedene Untersuchungen verweisen auf den Einfluss negativer Emotionen bei der Ablehnung von unfairen Angeboten im Ultimatum-Spiel. Van't Wout et al. (2006) konnten zeigen, dass Personen, die nach dem Empfang eines unfairen Angebots eine höhere Hautleitfähigkeit aufwiesen (ein Zeichen für erhöhte emotionale Erregung), diese Angebote eher ablehnten. Die Stärke und Hartnäckigkeit, mit der Emotionen wirken, konnten Bosman, Sonnemans und Zeelenberg

(2001) nachweisen. Sie verglichen zwei Gruppen: eine, in der sofort entschieden werden musste, ob ein Angebot im Ultimatum-Spiel angenommen wird oder nicht, und eine, die erst nach einer Stunde entscheiden musste. Die Personen, die Zeit hatten, um sich „abzukühlen“ („cooling off“), entschieden nicht anders als die Kontrollgruppe. Personen scheinen also nicht im Affekt zu handeln. Die Emotionen spielten offensichtlich dennoch eine Rolle, denn die „abgekühlten“ Versuchspersonen ärgerten sich sofort aufs Neue, wenn sie nach einer Stunde wieder mit dem ungerechten Angebot konfrontiert wurden. Auf die mögliche gegenläufige Wirkung von Kognitionen, die dafür sprechen zu akzeptieren, und Emotionen wie Ärger, die einen Entscheider motivieren ein unfaires Angebot abzulehnen, verweisen Sanfey et al. (2003).

#### *Variationen des Ultimatum-Spiels*

Spielt man das Ultimatum-Spiel mehrfach hintereinander, zeigt sich nur ein relativ geringer Effekt dieser Wiederholungen. Die Angebote und die Ablehnungsrate sinken leicht, was evtl. auch auf befriedigte Rachelüste zurückgeführt werden kann (Camerer, 2003). Es lässt sich insbesondere nicht beobachten, dass die Ergebnisse im wiederholten Ultimatum-Spiel – also mit steigender Erfahrung der Spieler – hin zum spieltheoretischen Gleichgewicht konvergieren (Roth, 1995).

Auch bei Veränderungen der Höhe der Einsätze zeigen sich keine grundlegend anderen Verhaltensmuster (Camerer & Fehr, 2004; Hoffman, McCabe & Smith, 1996).

Versuchspersonen reagieren recht sensibel darauf, ob Kommunikation möglich ist oder nicht. Wenn die Versuchspersonen bei einem Ultimatum-Spiel die Möglichkeit haben, miteinander zu kommunizieren (auch über das Spiel), kommt es in deutlich mehr Fällen zur Gleichverteilung des aufzuteilenden Betrages. Roth (1995) berichtet bei einem Experiment mit vollständiger Kommunikation einen Anteil der 50/50-Aufteilungen von 75 % verglichen mit 31 % ohne Kommunikation.



Die Versuchsperson-Versuchsleiter-Anonymität scheint eine eher geringe Rolle zu spielen. In einem wiederholten Ultimatum-Spiel mit und ohne Anonymität zwischen Versuchsleiter und Versuchsperson fanden Bolton und Zwick (1995) lediglich einen geringen Einfluss. Insbesondere zeigt sich, dass diese Form der Anonymität nicht generell zu Ergebnissen führt, die nahe oder gleich dem spieltheoretisch vorausgesagten Gleichgewicht sind (Roth, 1995).

Die Frage, ob eine Bestrafung möglich ist oder nicht, hat einen dramatischen Effekt. Bolton und Zwick (1995) verglichen ein klassisches Ultimatum-Spiel mit einem Impunity-Spiel (Straflosigkeit), in dem der Aufteilende ungestraft davon kam. Die Ablehnung der Aufteilung durch den Entscheider hat hier lediglich zur Folge, dass er nichts bekommt, der Aufteilende bekommt den Anteil, der ihm laut seiner eigenen Aufteilung zugedacht ist. Hier zeigte sich, dass keine noch so kleine Zuteilung abgelehnt wurde und die Ergebnisse sich deutlich dem spieltheoretischen Gleichgewicht annähern, in dem der Aufteilende (fast) alles bekommt und der Entscheider (fast) nichts.

Auf kulturelle Unterschiede bei Ultimatum-Spiel-Experimenten wird hier nur kurz eingegangen, da sie für die unten vorgestellten Untersuchungen nicht von Belang sind. Vergleicht man (entwickelte) Kulturen, zeigen sich zwar Unterschiede in den Mustern der Angebote und Ablehnungsquoten, diese sind jedoch nicht dramatisch (Camerer & Fehr, 2004). Das Verhalten von Stammeskulturen weicht allerdings teilweise deutlich von dem Verhalten in westlichen entwickelten Kulturen ab. Zwei (Haupt-)Variablen konnten identifiziert werden, die unterschiedliches Verhalten in den einzelnen Gesellschaften erklären können: der Grad der Marktintegration (findet regelmäßiger Marktaustausch zwischen den Individuen statt?) und die Wichtigkeit von Kooperation in einer Gesellschaft (die Erträge, die aus der Kooperation in einer Gesellschaft entstehen) (Henrich et al., 2005; Henrich et al., 2001). Die Ergebnisse deuten darauf hin, dass sowohl die Marktintegration als auch die Wichtigkeit der Kooperation in der

Gesellschaft positiv mit der Erwartung zusammenhängen, dass niedrige Angebote abgelehnt werden (Henrich et al., 2005; vgl. hierzu auch die empirische Untersuchung zur Interdependenz und indirekten Erträgen von Horne, 2004). Für einen detaillierten Überblick über kulturelle Unterschiede sei hier an andere Stellen verwiesen (Oosterbeek, Sloof & Van de Kuilen, 2004; Henrich et al., 2006; Henrich et al., 2001; Roth, 1995; Roth, Prasnikar, Okunofujiwara & Zamir, 1991; einen Überblick sowie Hinweise auf Probleme bei interkulturellen Vergleichen gibt Camerer, 2003).

Wie lassen sich die Unterschiede zwischen den theoretischen Voraussetzungen und den Ergebnissen der Experimente erklären? Die spieltheoretische Voraussage berücksichtigt ausschließlich das eigennützige Gewinnmaximierungsmotiv. Weder das Bedürfnis nach Fairness oder die Kraft der Emotionen bei der Entscheidung noch die Bereitschaft zur altruistischen Bestrafung werden berücksichtigt. Es kann aber der Einfluss verschiedener Motive sein, der zu einer bestimmten Aufteilung im Ultimatum-Spiel führt. Erstens kann sich der Aufteilende aus einem altruistischen Motiv heraus entscheiden, dem Entscheider einen angemessenen Teil des Kuchens zu geben. Zweitens kann es aber auch sein, dass der Aufteilende einen angemessenen Teil abgibt, da er befürchtet, dass der Entscheider die Aufteilung ablehnen könnte (Forsythe, Horowitz, Savin & Sefton, 1994; Roth, 1995; Harrison & McCabe, 1996; Fehr & Fischbacher, 2002). Es gibt also eine Verquickung von altruistischen und strategischen Beweggründen. Das Verhalten des Aufteilers kann also aus unterschiedlichen, nämlich altruistischen oder strategischen Motiven zum gleichen Verhalten führen (50/50-Aufteilung), während Menschen mit einem starken Bedürfnis für Gerechtigkeit in der Rolle des Entscheiders deutlich höhere Angebote ablehnen als strategisch-eigennützige Personen.

Ein Vergleich der Aufteilung in einem Ultimatum-Spiel mit derjenigen in einem Diktator-Spiel kann dazu genutzt werden, um die Motive voneinander zu trennen. (Camerer, 2003; Camerer & Fehr, 2004).

Ob und inwieweit der Aufteilende bei seiner Aufteilung die Bestrafungsneigung des Entscheiders vorhersieht und (strategisch) berücksichtigt, wird in den Studien 3 und 4 weiter unten experimentell untersucht.

#### **4.2.2. Diktator-Spiel (Dictator Game)**

Das Diktatorspiel modelliert eine Situation, in der ein (anonymer) Geldgeber eine Spende macht. Er kann damit keine Reputation aufbauen oder sonstige strategische Vorteile erhoffen. Auch dem Finder eines Geldbeutels ist mit Sicherheit klar, dass es eine einmalige Situation ist und dass die Person, die den Geldbeutel verloren hat, kein Veto einlegen kann, wenn er sich entschließt, den Geldbeutel nicht in einem Fundbüro abzugeben.

Das Diktator-Spiel wird in den Studien 3 und 4 als Verhaltensmaß für Altruismus eingesetzt und mit dem Ultimatum-Spiel verglichen. Mit Hilfe des Diktatorspiels kann gemessen werden, wie viel dessen, was die Aufteilenden im Ultimatum-Spiel anbieten, lediglich auf die Angst vor der Ablehnung des Angebots zurückzuführen ist (Camerer & Fehr, 2004).

Das Diktator-Spiel unterscheidet sich nur dadurch vom Ultimatumspiel, dass der Empfänger keine Möglichkeit zum Widerspruch hat. Das bedeutet, der Aufteilende entscheidet, welchen Betrag  $x$  er von einem Gesamtbetrag  $S$  an den Empfänger geben möchte. Die Aufteilung wird dann exakt so umgesetzt, wie es der Aufteilende – wie ein Diktator – beschlossen hat.

Zahlenbeispiele:

1. Der Aufteilende entscheidet, dass er von insgesamt 10 Euro 5 behält und 5 an den Empfänger abgibt. Beide erhalten entsprechend 5 Euro.

2. Der Aufteilende entscheidet, dass er von insgesamt 10 Euro 10 behält und 0 an den Empfänger abgibt. Der Aufteilende erhält entsprechend 10 Euro, der Empfänger nichts.

Dieses Spiel wird häufig als Messinstrument für Altruismus verwendet, denn in dieser Situation gibt es keine strategischen Überlegungen auf Seiten des Aufteilenden (Diktators). Da der Empfänger nicht ablehnen kann, gibt es keine Möglichkeit der Sanktionierung und auch aus den sonstigen Gegebenheiten (Anonymität, einmaliges Spiel) lässt sich kein anderer Grund für einen positiven Anteil für den Empfänger ableiten außer Altruismus.

Nach der Spieltheorie geben rationale, eigennutzorientierte Aufteilende im Diktator-Spiel Null Euro an den Empfänger. Es besteht keine Gefahr durch Ablehnung und jede abgegebene Geldeinheit reduziert den Nutzen des Aufteilenden.

Für die Aufteilung im Diktator-Spiel gelten aus psychologischer Sicht im Prinzip die gleichen Anmerkungen, die schon beim Ultimatum-Spiel gemacht wurden. Aus der zufälligen Zuordnung der Spieler zu den Positionen lässt sich nicht ableiten, dass einer der beiden Spieler mehr erhalten sollte. Anscheinend sehen das die Versuchspersonen auch so: Falk und Fischbacher (2006, S. 295 f.) fragten Versuchspersonen danach, welche Aufteilung „nett“ ist. Aufteilungen, die dem Empfänger weniger als 50% zusprachen, wurden als „nicht nett“ aufgefasst. Auch Fetchenhauer (unveröffentlichte Daten) erhielt auf die Frage nach der „moralisch richtigen“ Verteilung als Antwort die 50/50-Verteilung. Daraus lässt sich ableiten, dass die Gleichverteilung auch in diesem Fall der Referenz-Standard zur Beurteilung der Aufteilungen ist. Da das Spiel meist unter Anonymität stattfindet, stellt das Diktator-Spiel einen geeigneten Test für den reinen Altruismus einer Versuchsperson dar.

Die meisten Aufteilenden geben im Diktator-Spiel etwas an den Empfänger weiter. Die Zuteilungen liegen meist zwischen 10 % und 25 % der aufzuteilenden Summe. Die empirischen Ergebnisse deuten

also darauf hin, dass die Angst vor der Ablehnung im Ultimatum-Spiel eine wichtige Erklärung ist, aber eben nicht alles erklären kann (Camerer & Fehr, 2004; Forsythe et al., 1994). Hoffman, McCabe und Smith (1996) führen die (vom Standpunkt der Spieltheorie) zu hohen Angebote der Diktatoren auf die mangelnde Anonymität zwischen Diktator und Versuchsleiter zurück. Bolton und Zwick (1995) konnten in einem Versuch mit Versuchsleiter-Versuchsperson-Anonymität aber zeigen, dass der Einfluss der Anonymität zwar vorhanden, aber relativ gering ist. Trotz vollständiger Anonymität wurden positive Angebote gemacht und so gegen die Voraussagen der Spieltheorie verstoßen.

Fetchenhauer und Huang (2004) weisen einen Zusammenhang zwischen der Persönlichkeit der Versuchspersonen und ihrem Verteilungsverhalten nach. Bei der Untersuchung der Frage, ob eine Versuchsperson im Diktatorspiel eine 50/50-Aufteilung gewählt hat, zeigt sich, dass hohe Werte auf  $US_{\text{Opfer}}$  negativ mit der Entscheidung für eine 50/50-Verteilung zusammenhängen. Versuchspersonen, die sensibel auf eigene Benachteiligungen reagieren, teilen im Diktatorspiel also seltener 50/50 auf als Menschen mit niedrigen Werten auf dieser Skala. Das Gegenteil zeigte sich für Versuchspersonen mit hohen Werten auf  $US_{\text{Nutznießer}}$ .

#### *Variation des Diktator-Spiels*

Fehr und Fischbacher (2004b) nutzen ein Diktator-Spiel mit Bestrafung, um den Unterschied zwischen Second- und Third-Party-Punishment zu untersuchen. Das Diktator-Spiel mit Bestrafung unterscheidet sich vom einfachen Diktator-Spiel dadurch, dass der Empfänger in einer zweiten Stufe des Spiels die Gelegenheit zur Bestrafung erhält, allerdings unter Einsatz seines eigenen Geldes. Hierfür bekommt er 50 Punkte, damit er auch dann eine Bestrafung bezahlen kann, wenn der Aufteilende ihm nichts übertragen hat. Damit die Unterschiede, die durch die Aufteilung entstanden sind, aber konstant bleiben, erhält auch der Aufteilende auf der zweiten

Stufe 50 Punkte. Zieht der Empfänger dem Aufteilenden drei Punkte ab, kostet ihn das selbst einen Punkt.

Was sagt die Spieltheorie für das Diktator-Spiel mit Bestrafung voraus? Die Bestrafung kostet den Empfänger etwas und es entsteht ihm daraus kein (materieller) Nutzen. Daher wäre es für den Empfänger in einer einmaligen anonymen Interaktion rational und den Eigennutz maximierend, nicht zu bestrafen. Der Aufteilende weiß das und wird daher auf der ersten Stufe den kleinstmöglichen Betrag an den Empfänger überweisen.

Psychologisch sieht die Situation anders aus. Wenn der Verstoß gegen die Gleichverteilungsnorm als hinreichend gravierend wahrgenommen wird und die negativen Emotionen hinreichend stark sind, um eine Bestrafung zu motivieren, wird es zu einer Bestrafung auf der zweiten Stufe kommen. Wenn der Aufteilende diese Bestrafung antizipiert, wird er auf der ersten Stufe ein Angebot machen, das sich der Gleichverteilung annähert.

Die empirischen Ergebnisse stützen eher die psychologische als die spieltheoretische Betrachtungsweise (Fehr & Fischbacher, 2004b). Es wird gestraft und die Angebote sind weit von dem entfernt, was die Spieltheorie vorhersagen würde. Die Ergebnisse des Diktator-Spiels mit Bestrafung werden genauer im Kapitel „Vergleich von Second- und Third-Party-Punishment“ abgehandelt. Auf den Vergleich zwischen Ultimatum-Spiel und Diktator-Spiel und die daraus gezogenen Schlüsse geht Studie 3 genauer ein.

#### **4.2.3. Öffentliches-Gut-Spiel mit Bestrafung (public goods game)**

Ein Beispiel für ein öffentliches Gut, dessen Erstellung und Nutzung durch dieses Spiel modelliert wird, ist „saubere Luft“. Die mit Kosten verbundene Bereitstellung in Form von Maßnahmen zur Luftreinhaltung wird allerdings nur von manchen Nutzern geleistet. Andere, die Trittbrettfahrer (free-rider), tragen nicht zur Erstellung bei, können aber nicht von der Nutzung ausgeschlossen werden. Man kann

ihr Verhalten aber mit Strafen belegen. Die Feststellung einer Luftverschmutzung und die Strafverfolgung sind allerdings mit Kosten verbunden.

An einem Öffentlichen-Gut-Spiel (ÖGS) kann eine beliebige Zahl von Spielern teilnehmen. Alle sind zu Beginn des Spiels (bzw. jeder Runde) mit einem bestimmten Geldbetrag  $X$  ausgestattet. Die Spieler haben dann folgende Wahl: Sie können einen beliebigen Teil  $g$  ihres Geldes (zwischen 0 und  $X$ ) in ein öffentliches Gut investieren, den Restbetrag ( $X-g$ ), den sie nicht in das öffentliche Gut investieren, behalten sie. Die Beiträge aller Spieler zum öffentlichen Gut werden mit  $M(>1)$  multipliziert. Für gewöhnlich werden mehrere Runden gespielt. Nach jeder Runde hat jeder Spieler ein Einkommen, das aus seinem zurückbehaltenen Geld ( $X-g$ ) und den Einnahmen aus dem öffentlichen Gut besteht. Die Einnahmen berechnen sich aus der Summe der Beiträge der einzelnen Spieler multipliziert mit  $M$ . Sie werden gleichmäßig auf alle Spieler aufgeteilt (also auch auf diejenigen, die nichts beigetragen haben). Der persönliche Rückfluss einer Geldeinheit aus dem öffentlichen Gut beträgt  $M$  geteilt durch die Zahl der Spieler ( $N$ ), was kleiner als eins ist. Auf der anderen Seite geht jede Geldeinheit, die zurück gehalten wird, aber eins zu eins in das Einkommen ein. Daher ist es für jedes Individuum lukrativer, nichts zum öffentlichen Gut beizutragen. Insgesamt wären aber alle besser dran, wenn alle voll beitragen würden. In einer Gruppe rein eigennutzenorientierter Spieler würde also jeder  $X$  verdienen, da niemand etwas beiträgt. In einer Gruppe von Spielern, die alle voll in das Gut investieren, würde jeder  $M \cdot X$  verdienen, was mehr ist ( $M \cdot X = N \cdot X \cdot M / N$ ; alle Spieler ( $N$ ) tragen voll bei ( $X$ ) und die Investition wird mit  $M$  multipliziert; das Ergebnis wird auf alle Spieler ( $N$ ) aufgeteilt). Wenn aber ein Spieler nichts beiträgt und die anderen alles, dann verdient er noch mehr. Das individuelle und das soziale Optimum fallen also auseinander, und es besteht ein Anreiz, nichts beizutragen. (Camerer & Fehr, 2004; Fehr & Gächter, 2000). Das Öffentliche-Gut-Spiel wird ausschließlich unter Second-Party-Punishment abge-

handelt, da es definitionsgemäß keine unbeteiligte dritte Partei geben kann, ansonsten wäre es ein privates Gut (Bendor & Mookherjee, 1990).

Zahlenbeispiele:

1. Alle Spieler in einer Gruppe von vier Spielern entscheiden sich, ihre 20 Euro vollständig in das öffentliche Gut zu investieren. Die Verzinsung im öffentlichen Gut ist 1.6 (M). Die Auszahlung an alle Spieler beträgt dann jeweils 32 [ $0 + (4 \cdot 20 \cdot 1.6) / 4 = 0 + 32$ ].
2. Ein Spieler A, in einer Gruppe von vier Spielern, entscheidet sich, nichts von seinem ursprünglichen Geldbetrag von 20 Euro in das öffentliche Gut zu investieren. Die drei anderen Spieler investieren ihre 20 Euro vollständig in das öffentliche Gut. Die Verzinsung im öffentlichen Gut ist 1.6 (M). Die Auszahlung an Spieler A beträgt dann 44 [ $20 + (3 \cdot 20 \cdot 1.6) / 4 = 20 + 24$ ]. Die Auszahlung der anderen Spieler beträgt je 24 [ $0 + 24$ ].

In dem ursprünglichen ÖGS besteht die einzige Möglichkeit kooperativer Spieler, ihre unkooperativen Spielpartner zu bestrafen, darin, selbst nicht mehr zum öffentlichen Gut beizutragen. In einer Erweiterung des Spiels wird der ersten Stufe, auf der zum öffentlichen Gut beigetragen werden kann, eine zweite Stufe hinzugefügt, auf der die Spieler ihre Mitspieler mit monetären Strafen belegen können, wobei auch hier das Bestrafen mit Kosten für den Bestrafenden verbunden ist. Im Experiment von Fehr und Gächter kostete der Abzug von drei Geldeinheiten den Bestrafenden Spieler eine Geldeinheit (Fehr & Gächter, 2000)<sup>12</sup>.

Für beide Arten von Spielen – mit und ohne explizite Bestrafungsmöglichkeit – sagt die Spieltheorie unter der Annahme rationaler eigennutzorientierter Spieler das gleiche Ergebnis voraus: Keine Kooperation, d.h. kein Spieler trägt etwas zum öffentlichen Gut bei. Für den Fall ohne Bestrafung ist dies leicht nachzuvollziehen. Da das

---

<sup>12</sup> In den Spielen wird meist sichergestellt, dass es nicht zu negativen Auszahlungen für die Spieler kommen kann.



Individuum höhere Erträge erhält, wenn es nicht beiträgt, wird ein rationaler eigennutzinteressierter Spieler nichts beitragen. Diese Überlegung gilt für alle Spieler gleichermaßen, so dass nichts beigetragen wird. Für den Fall mit Bestrafung ist die Begründung umfangreicher. Da die Bestrafung den Bestrafenden etwas kostet, ihm aber keinen materiellen Nutzen bringt, wird kein rationaler eigennutzorientierter Spieler bereit sein zu bestrafen. Das gilt für alle Spieler und daher muss auch keiner eine Bestrafung befürchten. Aus diesem Grund verhalten sich auch alle so, als ob es keine Bestrafungsmöglichkeit gäbe, und tragen nichts bei (Fehr & Gächter, 2000).

Die psychologische Interpretation des ÖGS ist schwieriger als bei den vorangegangenen Spielen, da hier keine eindeutige Norm den Beitrag zum öffentlichen Gut vorschreibt. Aus der Equity-Theorie lässt sich aber ableiten, dass diejenigen, die (voll) zum öffentlichen Gut beitragen, ärgerlich auf diejenigen reagieren werden, die nicht beitragen, aber dennoch einen Ertrag aus dem öffentlichen Gut erhalten. Die einzige Möglichkeit, das Verhältnis zwischen Erträgen und Beiträgen im ÖGS ohne Bestrafung zu beeinflussen, besteht darin, die Beiträge zum Öffentlichen Gut zu reduzieren. Dieser Vorgang wird so lange dauern, bis das Verhältnis von Erträgen und Beiträgen für alle Spieler annähernd gleich ist. Wenn einige Spieler nie beitragen, kommt es zum Kollaps des öffentlichen Gutes, da dieser Zustand der einzige ist, in dem Erträge und Beiträge für alle Beteiligten gleich sind. Für das ÖGS mit Bestrafungsmöglichkeit weicht die psychologische Voraussage wieder deutlich von der spieltheoretischen ab. Den Spielern bietet sich neben der Einschränkung ihres Beitrags zum öffentlichen Gut nun die Möglichkeit, das Verhältnis der Erträge zu den Beiträgen im Vergleich zu den Trittbrettfahrern dadurch zu beeinflussen, dass deren Erträge durch die Bestrafung verringert werden. Als Motivation für die Bestrafung werden wiederum negative Emotionen angenommen.

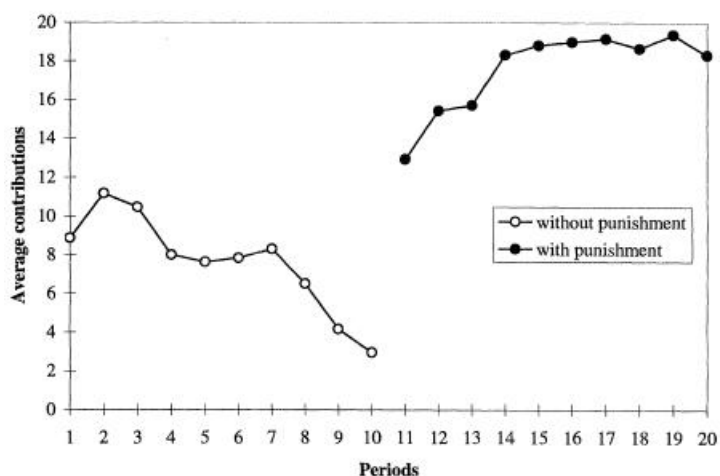
In einer Übersicht von Fehr und Schmidt (1999) über 12 Experimente mit ÖGS ohne Bestrafung zeigt sich das folgende Muster: Die Spieler starten auf einem mittleren Beitragsniveau (40 – 60 % ihrer Anfangssumme) und die Beiträge zum öffentlichen Gut sinken dann im Verlauf des Spiels, bis in der letzten Runde fast niemand mehr etwas beiträgt (73 % aller Spieler wählten einen Beitrag von 0). Die Kooperation kommt also zum Erliegen, alle Spieler verhalten sich wie Trittbrettfahrer.

Die Abnahme der Kooperation in den Experimenten wird mit unterschiedlichen Präferenzen der Versuchspersonen erklärt. Fischbacher, Gächter und Fehr (2001) erheben in einem einmaligen ÖGS ohne Bestrafungsmöglichkeit die Präferenzen der Versuchsteilnehmer mit Hilfe der Strategiemethode (vgl. Kapitel zur Strategiemethode weiter unten). Neben ca. 50% bedingt kooperativen Versuchsteilnehmern finden sie ca. 30% reine Trittbrettfahrer.

Bei wiederholten Spielen ist nach diesen Ergebnissen davon auszugehen, dass auch die bedingt kooperativen Teilnehmer mit der Zeit die Kooperation verweigern, wenn sie (frustriert) feststellen, dass die anderen unkooperativ sind (Andreoni, 1995). Sollten sie das Bedürfnis haben, die Unkooperativen zu bestrafen, ist die Verweigerung des eigenen Beitrags die einzige Möglichkeit im Spiel ohne (explizite) Bestrafung (Sigmund & Hauert, 2002). Deshalb kann schon ein geringer Anteil an Trittbrettfahrern ausreichen, um die Kooperation in einer Gruppe zum Erliegen zu bringen, wenn keine Bestrafung möglich ist (Fehr, 2004; Fehr & Fischbacher, 2002).

Die Ergebnisse des Spiels verändern sich durch die Einführung einer Bestrafungsmöglichkeit allerdings dramatisch. Können die Spieler Trittbrettfahrer bestrafen, zeigt sich ein viel höheres Kooperationsniveau. Es sinkt auch im Verlauf des Spiels nicht ab. Teilweise steigen die Beiträge bis zum Spielende sogar an. Im Allgemeinen zeigt sich, dass ein Spieler umso heftiger bestraft wird, je mehr er vom durchschnittlichen Beitrag der anderen Spieler abweicht. Einen wichtigen

Einfluss auf die Kooperation hat die Zuordnung der Spieler in den Spielrunden (das so genannte „matching protocol“): Wenn die Spieler jede Runde neue Spielpartner zugeordnet bekamen (Stranger-Bedingung), wurde zwar weit mehr in das öffentliche Gut einbezahlt als ohne Bestrafung (im Durchschnitt 58 % der Anfangsausstattung), allerdings stellte sich kein eindeutiges Gleichgewicht ein. In der letzten Runde waren die gezahlten Beiträge zum öffentlichen Gut breit über alle möglichen Beiträge verteilt. Ein Grund hierfür ist sicher in der Art der Zuordnung zu finden. Da die Spieler nicht davon ausgehen konnten, wieder mit dem gleichen Partner zusammen zu treffen, wurde strategische Bestrafung ausgeschlossen. Insgesamt war die Bestrafung in der Bedingung mit wechselnden Partnern auch geringer. In dem Fall, dass die Spieler alle Runden des Spiels mit den gleichen Spielpartnern zusammen blieben (Partner-Bedingung), konvergierten die Beiträge eindeutig hin zur vollen Kooperation. In der letzten Runde zahlte der Großteil der Spieler (82,5 %) den gesamten Betrag in das öffentliche Gut ein. Der Unterschied könnte größer nicht sein und soll hier für die Interaktion mit den gleichen Partnern noch einmal kurz zusammengefasst werden: In der Bedingung mit Bestrafung tragen die Spieler am Ende fast in voller Höhe zum öffentlichen Gut bei, besteht keine Bestrafungsmöglichkeit sinkt die Kooperation auf nahezu Null. Dieses Ergebnis, das mit den weiter oben gemachten psychologischen Voraussagen übereinstimmt, wird in Abbildung 1 sehr deutlich (Fehr & Gächter, 2000).



**Abbildung 1: Kooperation im Öffentliches-Gut-Spiel ohne und mit Bestrafung (Fehr & Gächter, 2000, S. 989)**

Verschiedene Autoren führen die anfängliche Kooperation im ÖGS ohne Bestrafung zum Teil auf Verwirrung zurück. Der Verfall der Kooperation wird dann mit dem steigenden Verständnis und weicherer Verwirrung über die Auszahlungen in späteren Runden erklärt (Houser & Kurzban, 2002; Andreoni, 1995). Camerer und Fehr (2004) widersprechen dieser Ansicht von Lerneffekten aus folgendem Grund: Bei einem erneuten Start des ÖGS wurde verschiedentlich gezeigt, dass die Probanden wieder höhere Beträge in das öffentliche Gut investieren, teilweise sogar wieder soviel wie zu Beginn des Experiments.

Ein interessanter Nebeneffekt dieses Versuchsaufbaus ist, dass die Kosten für die Bestrafenden immer geringer werden. Erstens steigt die Kooperation, sodass Bestrafung weniger häufig nötig ist und zweitens genügen in späteren Durchgängen die Androhung und die schmerzhafteste Erinnerung an die Bestrafung, um die potentiellen Trittbrettfahrer zu disziplinieren (Fehr & Fischbacher, 2002).

#### **4.2.4. Gefangen-Dilemma-Spiel (prisoners dilemma)**

Das klassische Gefangen-Dilemma beschreibt eine Situation in der zwei Häftlinge durch Kooperation bzw. Koordination ihrer Aussagen die Höhe ihrer Bestrafung durch einen Richter beeinflussen können

(vgl. Dawes, 1980). Leugnen beide, bekommen sie eine relativ geringe Strafe und können nur für kleinere nachweisbare Straftaten verurteilt werden; koordiniertes Handeln erbringt das beste Ergebnis für beide. Gestehen beide, bekommen sie selbstverständlich eine höhere Strafe, als wenn sie schweigen und man ihnen wenig nachweisen kann, aber die Geständnisse werden strafmildernd berücksichtigt. Gesteht einer und der andere leugnet, erhält der Geständige eine deutliche Strafmilderung und kommt sogar besser weg als wenn beide gestehen (Kronzeugenregelung, beste Lösung für einen Spieler), der Ungeständige erhält eine massive Bestrafung, da er durch das Geständnis seines Komplizen schwer belastet wird (schlechteste Lösung für den ungeständigen Spieler).

Im einfachsten Fall spielen zwei Spieler. Das Spiel lässt sich aber auf beliebig viele Spieler erweitern. Von der Struktur ähnelt das Gefangen-Dilemma-Spiel dem Öffentliches-Gut-Spiel. Die Spieler haben jeweils zwei Handlungsmöglichkeiten: Kooperieren oder Nicht-Kooperieren (defection). Die Erträge der beiden Spieler hängen jeweils von der eigenen und der Handlung des Spielpartners ab. Beide Spieler wählen ihre Handlung simultan, d.h. wenn Spieler 1 darüber entscheidet, ob er sich kooperativ verhalten will oder unkooperativ, weiß er nicht, wie Spieler 2 sich entschieden hat. Die verschiedenen Ergebnisse lassen sich in einer 2\*2-Matrix darstellen (Gefangenenspiel):

Spieler 2		Spieler 1	
S1	gesteht nicht	S2 gesteht nicht	S2 gesteht
		S2: 2 Jahre S1: 2 Jahre	S2: 1 Jahr S1: 8 Jahre
S1	gesteht	S2: 8 Jahre S1: 1 Jahre	S2: 5 Jahre S1: 5 Jahre

**Tabelle 1: Strafen im Gefangen-Dilemma (S1: Spieler 1; S2: Spieler 2)**

Zahlenbeispiele:

1. Spieler 1 gesteht nicht, Spieler 2 gesteht auch nicht. Beide können nur wegen kleinerer Vergehen verurteilt werden und erhalten beide 2 Jahre Haft.
2. Spieler 1 gesteht, Spieler 2 gesteht nicht. Spieler 1 erhält aufgrund der Kronzeugenregelung ein Jahr Haft, Spieler 2 hingegen 8 Jahre.

Meist wird das Gefangenen-Dilemma mit positiven Erträgen gespielt anstatt wie im Beispiel mit Haftstrafen. Die Logik bleibt jedoch exakt dieselbe. Für jeden einzelnen Spieler ist es gewinnmaximierend, nicht zu kooperieren, das soziale Maximum wird aber nur bei Kooperation erreicht.

Nach der Spieltheorie streben beide Spieler ihr privates Gewinnmaximum an (im Bsp. 1 Jahr Haft). Es ist für jeden Spieler einzeln betrachtet immer besser zu gestehen und zwar unabhängig von der Strategie des anderen. Daher werden beide Spieler gestehen und es wird sich ein für beide nicht optimales Ergebnis einstellen (5 Jahre Haft für beide).

Auch hier kommt die Sozialpsychologie zu anderen Vorhersagen als die Spieltheorie. Die Annahme, dass Menschen mehr Geld weniger Geld vorziehen (oder lieber kürzer im Gefängnis sind), ist sehr plausibel. Gleichzeitig sind sie aber auch in ein Geflecht von Normen und Regeln eingebunden, die die enthemmte Verfolgung eigennütziger Ziele bremsen. Internalisierte Normen der Kooperation sollten zumindest bei einem Teil der Versuchspersonen dazu führen, dass sie kooperieren und das soziale Optimum erreicht wird (Biel & Thøgersen, 2007; vgl. hierzu auch Frank, 2004, der die gleichzeitige Existenz von kooperativen und unkooperativen Individuen in seinem Modell annimmt).

Empirische Untersuchungen mit dem 2-Personen-Gefangenen-Dilemma zeigen, dass auch hier die spieltheoretische Vorhersage

nicht zutrifft. Die Anzahl der Spieler, die kooperieren, ist deutlich von Null verschieden. Im Durchschnitt kooperieren ca. 50% der Spieler (Camerer & Fehr, 2004). Wird den Spielern die Möglichkeit zu kommunizieren gegeben, steigt die Kooperationsrate stark an. Insbesondere mündliche Versprechen zu kooperieren sind offensichtlich alles andere als „cheap talk“ (Frank, 2004).

Falk, Fehr und Fischbacher (2005) verändern die Struktur des Spiels, indem sie eine zweite Stufe mit der Möglichkeit zur Bestrafung unter Kosten einführen. Nachdem alle Spieler auf der ersten Stufe simultan ihre Entscheidungen gefällt haben, werden die Entscheidungen bekannt gegeben. Auf der zweiten Stufe können sich die Spieler gegenseitig bestrafen. Diese Bestrafung ist für den Bestraften und den Bestrafenden mit Kosten verbunden.

Diese Variation ändert das vorhergesagte Verhalten lediglich für die sozialpsychologische Vorhersage. Zumindest diejenigen Personen, die eine Kooperationsnorm internalisiert haben, werden unkooperative Spieler auch unter eigenen Kosten bestrafen.

Es zeigen sich deutliche Unterschiede bei der Bestrafung durch kooperative und unkooperative Spieler. Kooperative Spieler bestrafen mehr und härter. Außerdem bestrafen kooperative Spieler (fast) ausschließlich unkooperative Spieler, während unkooperative Spieler sowohl kooperative (böartige Bestrafung) als auch unkooperative Spieler bestrafen.

Ein weiteres interessantes Muster ergibt sich bei der Variation der Sanktionshöhe, in dem der Einsatz von einem Punkt auf Seiten des Strafenden zu unterschiedlichen Abzügen beim Bestraften führt. In der Experimentalbedingung mit niedriger Bestrafung steht der zur Bestrafung eingesetzte Punktbetrag in einem 1:1-Verhältnis zum Abzug. Ein eingesetzter Bestrafungspunkt führt zu einem Punkt Abzug beim Bestraften. In der Bedingung mit hoher Bestrafung wird zudem zwischen der Strafe für Kooperative (3.3 Punkte) und der Strafe für Unkooperative (2.5 Punkte) unterschieden. Ein Punkt, der

zur Bestrafung eines Kooperativen eingesetzt wird, führt demnach zu einem Abzug beim Bestrafen von 3.3 Punkten. Ein Punkt, der zur Bestrafung eines Unkooperativen eingesetzt wird, führt zu einer Bestrafung von 2.5 Punkten. Die unterschiedliche Höhe der Bestrafung wird eingeführt, um die Hypothese zu prüfen, dass bei niedrigeren Kosten der Bestrafung mehr gestraft wird.

Für kooperative Spieler zeigt sich dieser Zusammenhang nicht, sie bewirken in beiden Bedingungen annähernd gleiche Abzüge bei den unkooperativen Mitspielern. Da diese Abzüge in der Bedingung mit niedriger Bestrafung (Verhältnis Kosten zu Bestrafung 1:1) aber deutlich mehr kosten, kann auf die Stärke ihrer Motive geschlossen werden. Diese sind offensichtlich stärker als Kosten-Nutzen-Berechnungen.

Für unkooperative Spieler zeigt sich folgendes: Sie bestrafen in der Bedingung mit hoher Bestrafung sowohl unkooperative Spieler (Verhältnis Kosten zu Bestrafung 1:2.5) als auch kooperative Spieler (Verhältnis Kosten zu Bestrafung 1:3.3). In der Bedingung mit der niedrigen Bestrafung (Verhältnis Kosten zu Bestrafung 1:1) strafen die unkooperativen Spieler nahezu gar nicht mehr. Diese Muster lassen sich mit unterschiedlichen Motiven der Bestrafung erklären. Während kooperative Spieler eher die Verletzung der Kooperationsnorm bestrafen, scheinen unkooperative Spieler daran interessiert zu sein, die Unterschiede der Erträge zwischen sich selbst und allen anderen Spielern zu vergrößern. Die Bestrafung der Verletzung der Kooperationsnorm (unkooperative Spieler) ist in beiden Bedingungen möglich, deshalb ziehen kooperative Spieler den unkooperativen Spielern auch in beiden Situationen annähernd gleich viel ab. Die Maximierung der Unterschiede in den Erträgen ist allerdings nur in der Bedingung mit der hohen Bestrafung möglich, da eine Bestrafung mit dem Verhältnis 1:1 (niedrige Bestrafung) die Unterschiede gleich lässt. Aus diesem Grund bestrafen bössartige unkooperative Spieler nur in der Bedingung mit der hohen Bestrafung.



Hierin könnte sich ein unterschiedliches Verständnis der Situation ausdrücken: Kooperative Spieler haben offensichtlich Gruppeninteressen im Auge, deshalb kooperieren sie und bestrafen Normbrecher. Unkooperative Spieler sehen evtl. eher den Wettbewerbsaspekt der Situation, kooperieren deshalb nicht und verstehen konsequenterweise den Bestrafungsmechanismus nicht als solchen, sondern als Instrument, die Unterschiede zwischen sich und den anderen zu erhöhen (Wettbewerbsmotiv).

#### **4.2.5. Vertrauens-Spiel (Trust Game) bzw. Investitions-Spiel**

Im Internet werden viele Geschäfte (mehr oder weniger) anonym abgewickelt. Es ist keine Seltenheit, dass diese Interaktionen einmalig sind und die beiden Geschäftspartner sich nicht kennen. Wenn ein Verkäufer die Ware bereits geschickt hat, muss er darauf vertrauen, dass der Käufer den (vollen) Geldbetrag auch überweist. Eine solche Situation kann mit dem Vertrauensspiel modelliert werden.

Am Vertrauens-Spiel (Berg, Dickhaut & McCabe, 1995) nehmen zwei Personen A und B teil. Person A erhält einen Betrag  $X$  vom Versuchsleiter und hat zwei Alternativen, mit dem Geld zu verfahren: 1. Sie behält den gesamten Betrag  $X$  für sich. In diesem Fall erhält Person B kein Geld und das Spiel ist beendet. 2. Sie investiert und gibt den Betrag  $X$  an Person B. In diesem Fall wird der Betrag vom Versuchsleiter mit einem Faktor  $r$  vervielfacht, so dass Person B dann  $r \cdot X$  erhält. Person B hat nun auch wieder zwei Optionen: 1. Sie behält  $r \cdot X$  für sich und gibt nichts an Person A zurück. Person A bekommt am Ende des Spiels dann nichts ausbezahlt, Person B erhält  $r \cdot X$  ausbezahlt und das Spiel ist beendet. 2. Person B verhält sich vertrauenswürdig und gibt einen Betrag  $Y$  zurück an Person A. Person A erhält am Ende des Spiels  $Y$  ausbezahlt, Person B  $r \cdot X - Y$  und das Spiel ist beendet. Wenn Person A also vertrauensvoll ihr Geld an Person B gibt und Person B sich als vertrauenswürdig erweist, kann sich das Vertrauen von Person A bezahlt machen. Behält Person B jedoch den gesamten Betrag  $r \cdot X$  für sich, wäre es für Person A besser gewe-

sen, X für sich zu behalten und nicht zu vertrauen. Der zweite Schritt – die mögliche Rückgabe eines Teils des Geldes durch Person B – gleicht dem Diktator-Spiel, da Person A keinen Einfluss nehmen kann.

Zahlenbeispiele:

1. Person A entscheidet sich, Person B zu vertrauen und gibt ihren gesamten Anfangsbetrag von 10 Euro an Person B weiter. Der Versuchsleiter verdreifacht den weitergegebenen Betrag. Person B bekommt also 30 Euro. Person B entscheidet sich im Gegenzug, 15 Euro an Person A zurückzugeben. Spieler A und Spieler B erhalten am Ende des Spiels beide 15 Euro.
2. Person A entscheidet sich, Person B nicht zu vertrauen und gibt nichts an Person B weiter. Das Spiel endet und Person A erhält 10 Euro, Person B erhält nichts.

Das Vertrauens-Spiel stellt ein sehr einfaches Modell einer Interaktion dar. Meist wird es nur einmal gespielt, die Spieler sind anonym und Kommunikation ist nicht erlaubt. Aus diesem Grund wurde dem Vertrauens-Spiel auch die Fähigkeit abgesprochen, wirkliches menschliches Vertrauen abzubilden. Camerer (2003) sieht in der einfachen Form aber sogar einen Vorteil, denn das Spiel misst dadurch Vertrauen in seiner reinen Form. Folgt man Luhmann (2000, S. 27), dass Vertrauen ein „Problem der riskanten Vorleistung“ ist, erweist sich das Vertrauens-Spiel als gutes Modell von Vertrauenssituationen, in denen zwei (oder mehr) Personen zum beiderseitigen Wohlergehen miteinander kooperieren können.

Was sagt die Spieltheorie für das Vertrauensspiel voraus? Da es nicht im Eigeninteresse der vollständig rational und eigennützig handelnden Person B liegt, etwas zurückzugeben, wird sie dies entsprechend der Theorie auch nicht tun. Die Vorhersage ist hier analog zum Diktator-Spiel. Person A kann dieses Verhalten vorhersehen, da sie weiß, dass sich Person B eigennützig verhalten wird. Aus diesem Grund wird Person A kein Geld an Person B geben und den gesamten Betrag

X für sich behalten. Da das Geld bei der Weitergabe vom Versuchsleiter vervielfacht wird, wäre ein besseres Ergebnis für beide Spieler möglich. Beide könnten zum Beispiel je die Hälfte von  $r \cdot X$  erhalten, was immer größer als  $X$  ist, wenn  $r$  größer ist als 2 (was im Versuch meist der Fall ist). Das kollektiv bessere Ergebnis wird so durch individuell rationales Verhalten verhindert (wie im bereits geschilderten Öffentlichen-Gut-Spiel und im Gefangenen-Dilemma).

Die psychologische Sichtweise auf das Vertrauensspiel ist eine andere. Person A muss davon ausgehen, dass Person B zwischen ihrem Eigennutz und Normen distributiver Gerechtigkeit abwägt. Verschiedene Gerechtigkeitsprinzipien (Gleichheitsprinzip, Leistungsprinzip) führen auch hier wieder zum gleichen Ergebnis: Es gibt keinen gerechten Grund, warum eine der beiden Personen mehr erhalten sollte als die andere. Aus der Gerechtigkeitsperspektive würde entsprechend folgen, dass Person B  $r \cdot X$  (hälftig) auf beide Spieler aufteilen sollte. Person A kann auf die Einhaltung dieser Gerechtigkeitsnormen allerdings lediglich vertrauen, diese aber nicht durchsetzen. Wie Person B tatsächlich aufteilt, hängt davon ab, ob sie ihren Eigennutz oder die Gerechtigkeitsaspekte stärker gewichtet (Fetchenhauer, Gresser & Haferkamp, in Druck).

Empirische Untersuchungen zeigen, dass Person A entgegen den Annahmen der Spieltheorie in der Regel Geld an Person B weitergibt (vgl. Berg et al., 1995). Allerdings lohnt sich das Vertrauen im Allgemeinen nicht, da zu wenig zurückgezahlt wird (vgl. Camerer & Fehr, 2004). Ein interessantes Ergebnis ist, dass die Weitergabe von Beträgen, die größer Null aber kleiner als der Maximalbetrag sind, offensichtlich nicht als Vertrauen gewertet wird (Pillutla, Malhotra & Murnighan, 2003). Die Weitergabe solcher Teilbeträge führt dann sogar zu schlechteren Endergebnissen für den vertrauensvollen Spieler A, als wenn alles oder gar nichts weiter gegeben wurde.

Fetchenhauer und Dunning (2005) konnten zeigen, dass sich die Entscheidung zu vertrauen deutlich von einer „normalen“ Entscheidung

unter Risiko unterscheidet. Wenn die Wahrscheinlichkeit, von Spielpartnern enttäuscht zu werden oder in einer Lotterie nichts zu gewinnen, gleich hoch ist, sind Versuchspersonen trotzdem eher bereit, vertrauensvoll Geld an einen Mitspieler weiterzugeben als an der Lotterie teilzunehmen. Außerdem erbrachten ihre Untersuchungen, dass die Rate der vertrauenswürdigen Personen im Allgemeinen unterschätzt wird. Sie fanden also, dass gemessen an der Wahrscheinlichkeit enttäuscht zu werden, das Vertrauen eher zu hoch ist, die allgemeine Einschätzung der Vertrauenswürdigkeit aber eher zu niedrig (Für eine wirtschaftspolitische Betrachtung dieser Fragen siehe Fetschenhauer et al., in Druck).

Die Frage, wovon das Vertrauen abhängt, wurde von Schlösser (2006) untersucht. Seine Versuchspersonen verhielten sich dann vertrauensvoll, wenn sie bei dem Gedanken daran, sich vertrauensvoll gegenüber ihrem Spielpartner zu verhalten, gut fühlten. Es war eher das „gute Gefühl“ im Moment der Weitergabe, das sie zur Weitergabe des Geldes veranlasste, als z.B. Kosten-Nutzen-Erwägungen.

Fehr, Fischbacher, von Rosenblatt, Schupp und Wagner (2002) zeigen, dass das Vertrauens-Spiel zur Messung von Vertrauen geeignet ist. Sie führten eine landesweite repräsentative Befragung durch, in der auch das Vertrauens-Spiel gespielt wurde. Es zeigte sich, dass diejenigen Personen, die angaben, in ihrem Leben Anderen häufig zu vertrauen, dies auch im Vertrauens-Spiel häufiger taten als Personen, die ganz allgemein ihrer sozialen Umwelt mit Skepsis begegnen.

#### **4.2.6. Moonlighting-Spiel**

Die Situation des Moonlighting-Spiels (Schwarzarbeit) ist der eines Schwarzarbeiters nachempfunden, der Zugang zu einer (schwarzen) Kasse hat, aus der er seine Materialien usw. bezahlen soll. Am Ende seiner Tätigkeit wird er bezahlt. Da die Tätigkeit aber insgesamt illegal ist, kann keiner der beiden Akteure Gelder oder Leistungen vom anderen gerichtlich einfordern, ohne selbst bestraft zu werden.

Eine gerichtliche Auseinandersetzung und Bestrafung des anderen wäre also für beide mit Kosten verbunden.

Mit Hilfe des Moonlighting-Spiels von Abbink, Irlenbusch und Renner (2000) ist es möglich, sowohl positive als auch negative Reziprozität (Bestrafung) zu untersuchen. Es spielen zwei Spieler A und B miteinander. Beide Spieler sind zu Beginn mit einem bestimmten Geldbetrag ausgestattet (12 Taler, die am Ende des Experiments in echtes Geld umgerechnet werden). Spieler A (der Schwarzarbeiter) kann nun von Spieler B entweder Geld nehmen oder ihm welches geben (jeweils bis zu 6 Taler). Wenn Spieler A Geld an Spieler B weiter gibt, erhält Spieler B den dreifachen Betrag dessen, was Spieler A weiter gegeben hat (also maximal 18 Taler). Im Gegenzug kann Spieler B nun entweder Geld an Spieler A zurückgeben (maximal 18 Taler) oder diesen bestrafen (mit maximal 6 Talern). Setzt Spieler B einen Taler für die Bestrafung ein, führt dies zu einem Abzug von 3 Talern bei Spieler A (Die Bestrafung und die Rückgabe von Geld mussten so vorgenommen werden, dass es nicht zu negativen Auszahlungen kommt).

Zahlenbeispiele:

1. Spieler A entscheidet sich, Spieler B 6 Taler zu geben. Spieler B bekommt die 6 Taler von Spieler A plus 12 Taler vom Versuchsleiter. Insgesamt erhält B also 18 Taler, den dreifachen Betrag dessen, was A ihm gegeben hat. Spieler B entscheidet sich im Gegenzug, Spieler A 9 Taler zurückzugeben. Spieler A erhält am Ende des Spiels also 15 Taler ( $6+9$ ) und Spieler B erhält 21 Taler ( $12+9$ ).
2. Spieler A entscheidet sich, Spieler B 3 Taler wegzunehmen. Spieler B entscheidet sich im Gegenzug, Spieler A mit 2 Talern zu bestrafen, die bei A zu einem Abzug von 6 Talern führen. Spieler A erhält am Ende des Spiels also 9 Taler ( $12+3-6$ ) und Spieler B erhält 7 Taler ( $12-3-2$ ).

Was sagt die Spieltheorie voraus? Spieler A kann damit rechnen, dass ein rationaler und eigennütziger Spieler B ihn nicht bestrafen

wird, weil es ihn etwas kostet. Außerdem kann Spieler A bei einem rational und eigennützigem Spieler B nicht davon ausgehen, dass er etwas zurückerhält, wenn er in Vorleistung tritt und Spieler B Geld gibt (bzw. seine Arbeitsleistung). Daher wird Spieler A dem Spieler B soviel Geld abnehmen wie möglich. Spieler B wiederum wird Spieler A nicht bestrafen, ihm aber auch kein Geld geben, unabhängig davon, wie jener sich vorher verhalten hat.

Die sozialpsychologische Betrachtung des Spiels lässt sich in weiten Teilen analog zu den anderen Spielen aufbauen. Auch in diesem Spiel bringt die Kooperation wieder Vorteile für beide Spieler. Ein kooperativer Spieler A kann wie im Vertrauensspiel durchaus damit rechnen, etwas zurück zu bekommen. In jedem Fall muss ein Spieler A, der B etwas wegnimmt, mit einer Strafe rechnen. Beide sind zu Beginn gleich ausgestattet und aus gerechtigkeitspsychologischer Sicht gibt es keinen Grund, hieran etwas zu ändern. Nimmt Spieler A dennoch etwas vom fairen Anteil des B, wird dieser verärgert reagieren und Spieler A wahrscheinlich bestrafen. Auch hier wird vermutlich wieder eine Abwägung stattfinden.

Die empirischen Ergebnisse entsprechen nicht den Voraussagen der Spieltheorie (Abbink et al., 2000). Sie lassen sich wie folgt beschreiben: Geld nehmen lohnt sich nicht, da es hart bestraft wird. Geld geben rentiert sich kaum, da es zu selten honoriert und manchmal ausgenutzt wird. Spieler A, die Spieler B Geld wegnehmen, werden in der Regel so hart bestraft, dass sich diese Strategie nicht lohnt, sie haben am Ende im Durchschnitt ungefähr die Hälfte ihrer Anfangsausstattung übrig. Spieler, die weder etwas nehmen noch etwas geben, erhalten am Ende im Durchschnitt ihre Anfangsausstattung wieder. Das gleiche Ergebnis zeigt sich für Spieler A, die etwas an Spieler B geben, auch sie haben am Ende im Durchschnitt ungefähr ihre Anfangsausstattung.

#### 4.2.7. Drei-Personen-Ultimatum-Spiel („Mixed Game“)

Eine reale Situation, die mit dem Drei-Personen-Ultimatum-Spiel modelliert werden kann, ist folgende: Ein Konsument erfährt, dass der Produzent eines von ihm gekauften Gutes seine Arbeiter, die sich dagegen nicht wehren können, sehr schlecht bezahlt. Der Konsument kann den Unternehmer bestrafen, in dem er das Gut nicht mehr kauft.

Das Drei-Personen-Ultimatum-Spiel wurde von Güth und van Damme (1998) entwickelt. Der Aufbau entspricht dem des Ultimatum-Spiels, erweitert um eine dritte passive Person (Dummy). Der Dummy hat keine Macht und kann keinen Einfluss auf die Entscheidung nehmen. Wieder steht ein Geldbetrag  $X$  zur Verteilung zur Verfügung, allerdings schlägt der Aufteilende in diesem Spiel eine Verteilung des Geldbetrags auf sich, den Entscheider und den Dummy vor. Der Entscheider entscheidet über die Gesamtverteilung, d.h. er entscheidet für den Dummy mit. Lehnt der Entscheider die Aufteilung ab, erhalten alle drei nichts. Im Prinzip spielt der Aufteilende ein Diktator-Spiel mit dem Dummy und ein Ultimatum-Spiel mit dem Entscheider (daher auch die Bezeichnung „Mixed Game“). Dieses Spiel wird auch in der weiter unten berichteten Studie 3 eingesetzt.

Zahlenbeispiele:

1. Der Aufteilende schlägt vor, dass der Betrag von 150 Euro gleichmäßig auf alle drei Personen aufgeteilt werden soll. Der Entscheider akzeptiert dies und alle drei Personen erhalten 50 Euro.
2. Der Aufteilende schlägt vor, dass der Betrag von 150 Euro wie folgt aufgeteilt werden soll: Aufteiler: 80, Entscheider: 40, Dummy: 30. Der Entscheider lehnt diesen Vorschlag ab und alle erhalten nichts.

Das Drei-Personen-Ultimatum-Spiel steht in gewisser Weise zwischen Second- und Third-Party-Punishment (vgl. weiter unten das Kapitel

zu Third-Party-Punishment). Im Bezug auf den eigenen Anteil handelt es sich bei einer Ablehnung durch den Entscheider um Second-Party-Punishment. Lehnt der Entscheider jedoch ab, weil ihm der Anteil für den Dummy zu gering erscheint – unabhängig von dem, was er selbst erhält – handelt es sich um Third-Party-Punishment. Mit diesem Spiel kann die interessante Frage untersucht werden, ob sich der Entscheider für die Zuteilung des Dummys interessiert oder ob er sich mit dem Aufteiler verbündet und sie gemeinsam den Dummy ausbeuten. Inwieweit Aufteilende in diesem Spiel mit Third-Party-Punishment rechnen und ob dies ihre Aufteilungsentscheidung beeinflusst, wird durch einen Vergleich mit dem Diktatorspiel möglich (vgl. unten Studie 3).

Die Spieltheorie sagt genau wie beim einfachen Diktator-Spiel voraus, dass dem Dummy nichts zugeteilt wird. Für den Entscheider wird das gleiche vorausgesagt wie im Ultimatumspiel, nämlich dass ihm die kleinstmögliche Geldmenge angeboten wird und er diese annimmt.

Grundsätzlich gelten internalisierte Normen distributiver Gerechtigkeit für alle Beteiligten einer Interaktion. Aus diesem Grund sagt die Gerechtigkeitspsychologie eine Gleichverteilung unter allen drei Spielern vorher. Aus gerechtigkeitspsychologischer Sicht spielt es zudem keine Rolle, dass der Dummy sich nicht wehren kann. Ganz im Gegenteil könnte diese Konstellation sogar die Wahrnehmung einer zu schützenden Person fördern, so dass auf die Belange des Dummy besonders geachtet wird. Bezüglich des Anteils für den Entscheider und für dessen Reaktion gelten aus sozialpsychologischer Sicht die gleichen Vorhersagen wie beim Ultimatum-Spiel. Aus den weiter oben gemachten Aussagen zum „disinterested punishment“ und der Tatsache, dass dafür die gleichen psychologischen Mechanismen unterstellt werden wie bei der Bestrafung für eigenes erlittenes Unrecht, kann abgeleitet werden, dass es auch bei ungerechten



Angeboten für den Dummy zur Ablehnung der Aufteilung kommen wird.

Die Ergebnisse empirischer Untersuchungen sind uneindeutig: Güth und Van Damme (1998) finden sehr geringe Angebote für den Dummy und auch sehr geringe Ablehnungsquoten für diese ungerechten Angebote. Hieraus schließen sie auf eine geringe intrinsische Fairnessmotivation. Außerdem stellen sie fest, dass die Probanden auf unterschiedliche Informationsbedingungen strategisch reagieren: Den Versuchspersonen wurden entweder alle angebotenen Anteile mitgeteilt (Aufteilender, Entscheider, Dummy) oder nur Teilinformationen gegeben (Dummy bzw. Entscheider). Wurde der Entscheider lediglich über den Anteil des Dummys informiert, war das Angebot des Anbieters an den Entscheider (und den Dummy) nahe Null. Betraf die Information den Entscheider selbst (Aufteilender, Entscheider, Dummy bzw. Entscheider), waren die Angebote höher, und umfassten ungefähr ein Drittel des Gesamtbetrages. Insgesamt wurden nur etwa 7% der Angebote abgelehnt.

Diese Ergebnisse unterscheiden sich deutlich von denen von Fetschenhauer und Huang (2004). Hier lehnen über 50% der Probanden eine Aufteilung ab, die dem Dummy weniger als seinen gerechten Anteil von einem Drittel zuspricht. Zudem zeigte sich auch hier wieder ein Einfluss der Ungerechtigkeitssensibilität. Es waren die Personen mit hohen Werten auf  $US_{\text{Nutznießer}}$  und  $US_{\text{Beobachter}}$ , die unfaire Verteilungen zu ihren eigenen und zu Ungunsten des Dummys ablehnten. Der erwartete Zusammenhang zwischen  $US_{\text{Opfer}}$  und der Ablehnung von unfairen Angeboten an die Entscheider selbst zeigte sich allerdings nicht.

Die Ergebnisse lassen sich jedoch nur schwer vergleichen, da Güth und van Damme zwischen verschiedenen Informationsbedingungen unterschieden und die Versuchspersonen teilweise mehrere Versuche durchliefen.

#### 4.2.8. Power-To-Take-Spiel

Die Besteuerung von Arbeitseinkommen und die Reaktion auf diesen Eingriff in den Besitz ist eine Situation, die sich mit dem Power-To-Take-Spiel modellieren lässt. Der Arbeitnehmer (Responder) kann mit unterschiedlichem Arbeitseinsatz auf verschiedene Steuersätze des Staates (Nehmer) reagieren. Zu hohe Steuern können dazu führen, dass der Arbeitnehmer nicht mehr so viel arbeitet, wie er es bei einem angemessenen Steuersatz tun würde. Damit „zerstört“ er einen Teil des Einkommens (vor Steuern) und entzieht es so dem Zugriff des Staates.

Im Power-To-Take-Spiel (PTTS) (alle weiteren Ausführungen zum PTTS beziehen sich auf: Bosman & van Winden, 2002; Bosman et al., 2005) spielen zwei Personen: Spieler 1 (Take-Authority, ab jetzt „Nehmer“) und Spieler 2 (Responder). Beide Personen arbeiten vor dem Spiel an einer Aufgabe, um ihr Einkommen für das Spiel ( $Y_{\text{Nehmer}}$  bzw.  $Y_{\text{Responder}}$ ) zu verdienen. Das eigentliche Spiel besteht aus zwei Stufen. Auf der ersten Stufe bestimmt der Nehmer die Nehmerrate, d.h. wie viel Prozent ( $t$ ) des Einkommens des Responders er nehmen wird ( $0 < t < 1$ ). Auf der zweiten Stufe des Spiels bestimmt der Responder einen Teil  $d$  ( $0 < d < 1$ ) seines Einkommens, den er zerstören möchte, bevor der Nehmer darauf zugreifen kann. Der Gesamtertrag des Nehmers besteht aus seinem Einkommen ( $Y_{\text{Nehmer}}$ ) plus dem Teil des Einkommens, den er vom Einkommen des Responders nach der Zerstörung  $(1-d)$  wegnimmt ( $t \cdot [1-d] \cdot Y_{\text{Responder}}$ ). Der Responder behält den Rest, der nach Zerstörung  $(1-d)$  und Zugriff  $(1-t)$  durch den Nehmer übrig bleibt ( $[1-t] \cdot [1-d] \cdot Y_{\text{Responder}}$ ). Wenn der Nehmer nichts nimmt ( $t=0$ ) und der Responder nichts zerstört ( $d=0$ ), erhalten beide genau den Betrag, den sie vorher verdient haben ( $Y_{\text{Nehmer}}$  bzw.  $Y_{\text{Responder}}$ ).

Der Responder kann den Nehmer im PTTS dadurch bestrafen, dass er Teile seines Einkommens zerstört. Hier wird direkt deutlich, dass die Bestrafung mit Kosten verbunden ist. Am prozentualen Anteil des

Nehmers ändert der Responder nichts, er kann lediglich die absoluten Beträge für den Nehmer und sich reduzieren.

Zahlenbeispiele:

1. Nehmer und Responder haben beide 10 Euro verdient. Der Nehmer beschließt, dem Responder 25% seines Einkommens wegzunehmen. Der Responder beschließt 20% seines Einkommens zu zerstören. Der Nehmer erhält 12 Euro: 10 Euro, die er selbst verdient hat plus 2 Euro vom Responder (25% von 8 Euro). Der Responder erhält 6 Euro, da er von den 10 Euro 2 Euro zerstört und der Nehmer ihm von den restlichen 8 Euro 2 Euro weggenommen hat.
2. Nehmer und Responder haben beide 10 Euro verdient. Der Nehmer beschließt, dem Responder 90% seines Einkommens wegzunehmen. Der Responder beschließt, 100% seines Einkommens zu zerstören. Der Nehmer erhält die 10 Euro, die er selbst verdient hat. Der Responder erhält nichts, da er sein Einkommen vollständig zerstört hat.

Die Voraussage der Spieltheorie ist wie immer recht einfach: Verlangt der Nehmer auf der ersten Stufe einen Anteil, der kleiner als 100% ist, kommt es auf der zweiten Stufe nie zur Bestrafung, da sich der Responder damit auch selbst schadet. Verlangt der Nehmer 100%, ist der Responder im Prinzip indifferent zwischen „Zerstörung“ und „keine Zerstörung“, beides bringt ihm keinen materiellen Nutzen.

Die sozialpsychologische Voraussage beruht wieder auf der gerechtigkeitspsychologischen Annahme, dass beiden Spielern grundsätzlich das gleiche zusteht. Insofern wäre die Voraussage, dass mit steigender Nehmerrate die negativen Reaktionen der Spieler B zunehmen. Es gibt allerdings auch Fälle, in denen es eine akzeptierte Nehmerrate gibt (z.B. akzeptieren die meisten Menschen einen positiven Steuersatz). In einem solchen Fall ist davon auszugehen, dass lediglich ungerechte Abweichungen hiervon bestraft werden.

Was zeigen die empirischen Befunde? Die spieltheoretischen Voraussetzungen werden nicht gestützt. 1. nehmen die Nehmer nicht das Maximum vom Responder, sondern durchschnittlich ca. 60 %. Auf der anderen Seite zerstören die Responder in Abhängigkeit von der Nehmerrate (t). Interessanterweise zerstören die Responder tendenziell entweder alles oder nichts. Ein weiterer interessanter Befund ist, dass die Zerstörungsrate (d) von negativen Emotionen abhängt, die ihrerseits wieder von der Nehmerrate (t) abhängen. Die Tatsache, dass entweder alles oder nichts zerstört wird, spricht gegen eine Abwägung von Kosten und Nutzen der Zerstörung bei diesen Versuchspersonen.

#### **4.3. Third-Party-Punishment: Bestrafung durch einen unbeteiligten Dritten**

In den Spielen des vorhergehenden Abschnitts waren die bestrafenden Personen selbst direkt von der unfairen Handlung betroffen. In den folgenden Spielen ist dies nicht der Fall. Jetzt werden Spiele betrachtet, in denen die bestrafende Person eine unbeteiligte dritte Person ist. Diese Person ist nicht Teil der ökonomischen Transaktion (bzw. Interaktion) zwischen den beiden Protagonisten, sie ist lediglich durch soziale Verbindungen oder allgemeine Verhaltensregeln der Gesellschaft mit ihnen verbunden. Hieraus leitet sich auch der Name Third-Party-Punishment (im Weiteren abgekürzt durch TPP) (Fehr & Fischbacher, 2004b) bzw. Third-Party-Sanctions (Bendor & Mookherjee, 1990) ab.

Die höchst spannende und folgenreiche Frage, die sich hier stellt, ist: Bestrafen Personen auch dann, wenn sie nicht selbst betroffen sind und diese Bestrafung sie etwas kostet? Die Antwort lautet: Ja, einige Personen bestrafen auch dann. Dieser Befund stellt eine große Herausforderung für die Eigennutz-Thesen der Spieltheorie dar. Da die meisten Experimente in diesem Bereich einmalig und anonym durchgeführt werden, kann Eigennutz nie die Triebfeder für Third-Party-Punishment sein (Fehr & Fischbacher, 2004a; Bendor & Mook-

herjee, 1990, verweisen darauf, dass strategische Reziprozität als Begründung wegfällt). Aus diesem Grund ist Bestrafung durch Dritte auch ein besserer Indikator für eine soziale Norm, da es sich von den Motiven deutlich von Second-Party-Punishment abhebt (Bendor & Mookherjee, 1990).

#### **4.3.1. Drei-Personen-Bestrafungs-Spiel**

Stellen wir uns folgende Situation vor: Ungekündigte Mitarbeiter profitieren von der Entlassung anderer, wenn ein Unternehmen dadurch seine Profitabilität erhöhen kann (ihre Arbeitsplätze sind dann sicherer). Dennoch solidarisieren sie sich unter Umständen mit ihren (Ex-) Kollegen und setzen sich für diese ein bzw. „bestrafen“ die als unfair wahrgenommene Firma durch schlechtere Arbeitsleistung oder ähnliches (vgl. Skarlicki & Kulik, 2005).

Das Drei-Personen-Bestrafungs-Spiel (Third-Person-Punishment-Game) geht auf Fehr und Fischbacher (2004b) zurück. Es ist dem Drei-Personen-Ultimatum-Spiel (s. o. Güth & van Damme, 1998) ähnlich. Allerdings wird hier dem Diktator-Spiel eine dritte Person hinzugefügt, die bestrafend eingreifen kann. Der Aufteiler teilt einen Geldbetrag  $X$  zwischen sich und dem Empfänger auf. Wie im Diktator-Spiel kann der Empfänger keinerlei Einfluss auf die Aufteilung nehmen. Die dritte Person ist hier – im Gegensatz zum Spiel von Güth und van Damme – nicht mehr direkt von der Aufteilungentscheidung des Aufteilers betroffen. Sie erhält unabhängig von der Aufteilung des Geldbetrags eigene Ressourcen vom Versuchsleiter. Nachdem die Aufteilung vollzogen wurde, hat die dritte Person die Möglichkeit den Aufteiler zu bestrafen, indem sie dessen Auszahlungen verringert. Diese Bestrafung kostet die dritte Person eigene Ressourcen, so dass mit diesem Spiel die reinste Form der altruistischen Bestrafung gemessen werden kann. Die dritte Person hat keinen (materiellen) Nutzen aus der Bestrafung, muss aber die Kosten der Durchsetzung der Verteilungsnorm tragen. Die Bestrafung wird modelliert, in dem die dritte Person dem Aufteiler Strafpunkte

auferlegen kann, wofür sie aber selbst Punkte einsetzen muss. Für jeden Punkt, den die dritte Person in die Bestrafung investiert, werden dem Aufteiler drei Punkte abgezogen.

Zahlenbeispiele:

1. Der Aufteiler teilt 100 Euro zwischen sich und dem Empfänger auf, indem er 50 Euro an den Empfänger gibt und 50 für sich behält. Die Dritte Person greift mit ihren zusätzlichen 50 Euro nicht in die Aufteilung ein. Alle drei Personen erhalten am Ende des Spiels 50 Euro.
2. Der Aufteiler teilt 100 Euro zwischen sich und dem Empfänger auf, indem er 20 Euro an den Empfänger gibt und 80 für sich behält. Die dritte Person hat unabhängig von den 100 Euro weitere 50 Euro, um in die Verteilung einzugreifen. Sie zieht dem Aufteiler 30 Euro ab und muss dafür 10 Euro bezahlen. Der Aufteiler erhält am Ende des Spiels 50 Euro (80-30), der Empfänger erhält 20 Euro und die dritte Person erhält 40 Euro (50-10).

Nach der Spieltheorie weiß der Diktator, dass eine rationale eigennützige dritte Person nie bestrafen würde, wenn sie selbst dafür etwas bezahlen muss und sie keinen Nutzen daraus ziehen kann. Aus diesem Grund verhält er sich so, wie im einfachen Diktator-Spiel und behält alles für sich.

Die sozialpsychologische Argumentation unterscheidet sich auch bei diesem Spiel deutlich von der spieltheoretischen: Es gelten die gleichen gerechtigkeitspsychologischen Annahmen bezüglich der Aufteilung und der Bestrafung wie im Drei-Personen-Bestrafungs-Spiel. Das Equity-Prinzip und das Gleichheitsprinzip schreiben eine Gleichverteilung vor. Nimmt der Aufteiler eine Aufteilung vor, die von der Gleichverteilung abweicht, wird aus sozialpsychologischer Perspektive eine Bestrafung durch die Dritte Person vorausgesagt.

Fehr und Fischbacher (2004b) finden in einem Versuch mit 66 Versuchspersonen (22 in der Rolle der dritten Partei) eine hohe Bestrafungsbereitschaft. Ein großer Teil der dritten Parteien (ungefähr 60%) bestraft den Aufteiler, wenn er dem Empfänger weniger als 50 Punkte von den insgesamt 100 Punkten anbietet. Je weniger der Aufteiler anbietet, desto stärker ist die Bestrafung. Wenn er alles für sich behält, wird er im Durchschnitt mit 14 Punkten bestraft, d.h. ihm werden 42 Punkte abgezogen. Eigennütziges Verhalten hat sich in diesem Versuch allerdings trotz der Bestrafung gelohnt. Aufteilende, die alles für sich behielten, verdienten im Schnitt 58 Punkte. Aufteilende, die 50/50 aufteilten, erhielten am Ende weniger. Sie erhielten sogar im Schnitt etwas weniger als 50 Punkte, da es auch zu einer kleinen Anzahl (nicht erklärter) Bestrafungen von 50/50-Aufteilungen kam.

Neben den tatsächlichen Aufteilungen und Reaktionen der dritten Parteien wurden auch die Erwartungen der Empfänger bezüglich der Bestrafung erhoben. Die Empfänger erwarteten von der dritten Partei, dass sie ungleich Aufteilende bestraft. Dies bekräftigt die Annahme über die saliente Verteilungsnorm. Die Vorstellungen der Empfänger bezüglich der Bestrafung waren erstaunlich akkurat. Für geringe Transfers lagen sie zwar über der tatsächlichen Bestrafung, dieser Unterschied war jedoch nicht statistisch signifikant. Dritte Parteien bestraften also ungerechte Aufteilungen und die Dummys erwarten dies von ihnen.

Die interessante Frage, ob die Aufteiler auch mit einer Bestrafung rechnen und dies ihr Verhalten lenkt, wird im empirischen Teil dieser Arbeit (Studie 3 und 4) beantwortet.

#### **4.3.2. Gefangenen-Dilemma mit Bestrafung durch Dritte**

Ein Gefangenen-Dilemma mit Bestrafung durch Dritte untersucht die Bestrafungsbereitschaft bei Verletzung der Kooperationsnorm. Der Aufbau des Gefangenen-Dilemmas auf der ersten Spielstufe ist identisch mit dem oben geschilderten bei Bestrafung durch die Mit-

spieler (second-party-punishment), nur dass jetzt keine Option der Bestrafung durch die Mitspieler besteht, sondern durch einen unteiligten dritten Spieler C. Auf der zweiten, der Bestrafungsstufe, kann nun diese Person C sowohl Spieler A als auch Spieler B aus dem Gefangenen-Dilemma bestrafen. Eine Bestrafung in Höhe von 3 Punkten bei einem der Spieler kostet Person C einen Punkt. Das Spiel ist wie folgt aufgebaut (Fehr & Fischbacher, 2004b):

Stufe 1: Spieler A und B besitzen jeweils 10 Punkte. Beide können ihre Punkte dem Spielpartner geben, dann werden sie vom Spielleiter verdreifacht. Wenn A seine 10 Punkte an B gibt, B seine Punkte jedoch behält, verdient A nichts und B verdient 40. Das soziale Optimum ist also eine Situation, in der beide Spieler 30 Punkte erhalten. Für jeden Spieler individuell ist es allerdings optimal, die 10 Punkte für sich zu behalten, so dass sich das spieltheoretische Gleichgewicht von 10 Punkten für beide Spieler ergibt.

Stufe 2 (Bestrafung): Spieler C beobachtet das Verhalten auf der ersten Stufe und erhält 40 Punkte, die er zum Abzug von Punkten bei A und B einsetzen kann (bis zu 20 Punkte von jedem der beiden Spieler). Spieler A und Spieler B erhalten je 15 Punkte zusätzlich zu den Ergebnissen aus der ersten Stufe des Spiels. Alle Spieler wussten über die Punkteausstattung der anderen, die Regeln des Spiels und die Bestrafungsoption Bescheid.

Zahlenbeispiele:

1. Spieler A gibt seine 10 Punkte weiter. Spieler B gibt seine 10 Punkte weiter. Die weitergegebenen Punkte werden vom Versuchsleiter verdreifacht. Beide Spieler besitzen nach dieser Runde also 30 Punkte. In der zweiten Runde erhalten Spieler A und B zusätzlich 15 Punkte und Spieler C erhält 40 Punkte. Spieler C bestraft nicht. Spieler A und B erhalten am Ende des Spiels je 45 Punkte (30+15) und Spieler C erhält 40 Punkte.
2. Spieler A gibt seine 10 Punkte weiter. Spieler B gibt seine 10 Punkte nicht weiter. Die weitergegebenen Punkte werden vom



Versuchsleiter verdreifacht. Spieler A besitzt nach dieser Runde 0 Punkte, Spieler B besitzt 40 Punkte ( $10+30$ ). In der zweiten Runde erhalten Spieler A und B zusätzlich 15 Punkte und Spieler C erhält 40 Punkte. Spieler C bestraft Spieler B mit 20 Punkten. Spieler A hat am Ende des Spiels 15 Punkte, Spieler B hat 35 ( $40+15-20$ ) und Spieler C hat 20 Punkte ( $40-20$ ).

Die spieltheoretische Voraussage in diesem Zusammenhang ist einfach: Es kommt zu keiner Bestrafung auf der zweiten Stufe, da Spieler C keinen persönlichen Nutzen aus der Bestrafung ziehen kann. Die Spieler A und B der ersten Stufe sehen das voraus und verhalten sich stets unkooperativ, da das die dominante Strategie für sie ist.

Wie auch schon in den anderen Spielen, wird aus sozialpsychologischer Sicht eine Bestrafung vorausgesagt, wenn eine Kooperationsnorm verletzt wird bzw. ein Spieler sich durch unkooperatives Verhalten auf Kosten anderer einen ungerechten Vorteil verschafft. Bezüglich der Kooperation auf der ersten Stufe gelten die oben gemachten Aussagen zu Normen der Kooperation im Gefangenen-Dilemma ohne Bestrafung durch Dritte.

Fehr und Fischbacher (2004b) berichten Ergebnisse, nach denen sich die meisten Menschen im Gefangenen-Dilemma bedingt kooperativ verhalten. Das bedeutet, dass sie nur kooperieren, wenn sie hinreichend sicher davon ausgehen können, dass ihr Spielpartner dies auch tut. Vor diesem Hintergrund stellt unkooperatives Verhalten bei Kooperation des Spielpartners eine starke Normverletzung dar; stärker als unkooperatives Verhalten, wenn beide unkooperativ handeln. Aus diesem Grund ist zu erwarten, dass 1. unkooperatives Verhalten allgemein bestraft wird und dass 2. ein unkooperativer Spieler bei einem kooperativen Mitspieler härter bestraft wird.

Die empirischen Ergebnisse bestätigen diese Vermutung. In beiden Fällen kommt es zu erheblicher Bestrafung durch die unbeteiligten Dritten. Spieler, die sich einem kooperativen Partner gegenüber unkooperativ verhielten, wurden von 45,8% der dritten Parteien be-

straft. Spieler, die sich einem unkooperativen Partner gegenüber unkooperativ verhielten, wurden immerhin noch von 20,8% der dritten Parteien bestraft (Es kam auch hier zu einer Bestrafung kooperativer Spieler, allerdings nur von einem kleinen Teil der Unbeteiligten und dann auch nur mit wenigen Punkten) (Fehr & Fischbacher, 2004a, 2004b).

#### **4.4. Vergleich von Second- und Third-Party-Punishment**

Inhaltlich besteht der Hauptunterschied darin, dass die bestrafende dritte Partei nicht von der ungleichen Aufteilungsentscheidung betroffen ist. Insbesondere in einmaligen Spielen gibt es für eine solche dritte Person also nie einen rationalen eigennützigen Grund, in die Aufteilungsentscheidung einzugreifen (Fehr & Fischbacher, 2004a). Das gleiche gilt natürlich auch für wiederholte Spiele, wenn die dritte Person sicher sein kann, nie wieder mit den bestraften Spielern zusammen zu spielen. Das bedeutet, dass Belohnungen und Bestrafungen einer dritten Partei auch nicht durch Eigennutz erklärt werden können. Aus diesem Grund sprechen Fehr und Fischbacher (2004b, S. 65) davon, dass diese Belohnungen und Bestrafungen die wahren normativen Standards des Verhaltens enthüllen („the rewards and sanctions of third parties reveal the truly normative standards of behavior“).

Camerer (2003) vermutet (siehe auch Fehr & Fischbacher, 2004b), dass bezüglich der Höhe der Bestrafung zu erwarten ist, dass Third-Party-Punishment geringer ausfällt als Second-Party-Punishment. Er geht davon aus, dass unterschiedliche Emotionen entstehen, wenn ein Normbruch am eigenen Leib erlebt oder lediglich beobachtet wird.<sup>13</sup> Beim Beobachten entsteht Empörung, die in der Intensität schwächer ist und eine geringere Motivation zur Bestrafung aufweist als der Ärger, der beim eigenen Erleben einer Ungerechtigkeit gefühlt

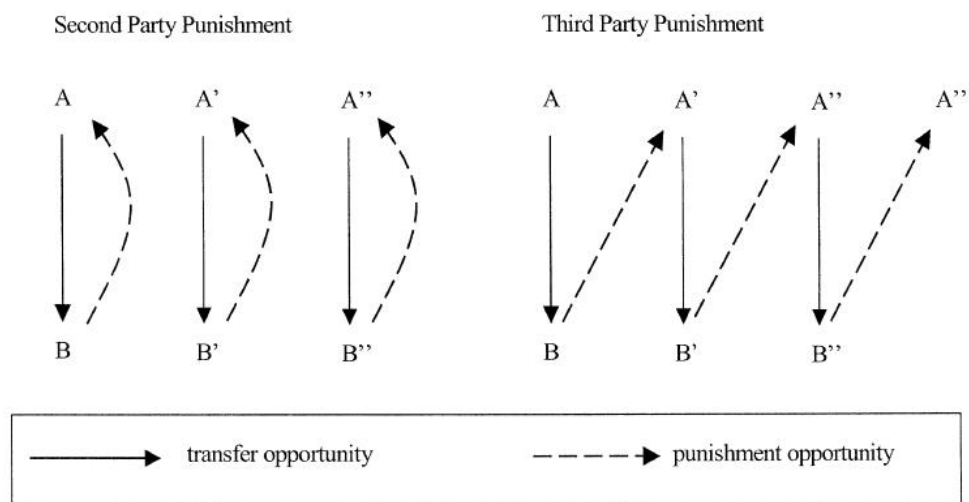
---

<sup>13</sup> Dies spiegelt sich auch in der Formulierung der Items der Ungerechtigkeitssensibilität wider (Schmitt et al., 1997).

wird. Ebenso deuten die Ergebnisse bezüglich der Ungerechtigkeits-sensibilität von Fetschenhauer und Huang (2004) in diese Richtung: deren Versuchspersonen reagierten sensibler, wenn sie selbst Opfer von Ungerechtigkeit waren, als wenn es andere betraf.

In einem direkten Vergleich haben Fehr und Fischbacher (2004b) die Unterschiede von Second- und Third-Party-Punishment beim Bruch von Verteilungs- und Kooperationsnormen untersucht:

Für die Untersuchung des Unterschieds bei Verteilungsnormen spielten mehrere Gruppen ein Diktator-Spiel. Zu Beginn des Experiments wurden alle Versuchspersonen zufällig einer der beiden Rollen (Diktator, Empfänger) zugeordnet. Jeweils ein Diktator und ein Empfänger bildeten eine Gruppe. Der Diktator A einer Gruppe gab an, wie viel er dem Empfänger B seiner Gruppe von 100 Punkten zukommen lassen wollte. Im Fall des Second-Party-Punishment konnte der Empfänger B „seinen“ Diktator A im Anschluss an die Aufteilung in einer zweiten Spielstufe bestrafen. Hierfür bekam Spieler B zu Beginn der zweiten Stufe 50 Punkte. Im Fall des Third-Party-Punishment konnte ein Empfänger B immer nur einen Diktator A´ einer anderen Gruppe bestrafen. B ist eine unbeteiligte dritte Partei, da der „fremde“ Diktator A´ in der Vergangenheit keinen Einfluss auf die Erträge von B hatte (vgl. Abbildung 2). Außerdem wurde jegliche Möglichkeit der strategischen Interaktion oder Reziprozität durch den Versuchsaufbau ausgeschlossen. Da der Aufbau der beiden Bedingungen so parallel wie möglich ist, kann auf diese Weise die Stärke des Second-Party-Punishment mit der Stärke des Third-Party-Punishment verglichen werden.



**Abbildung 2: Aufbau der Bestrafung beim Vergleich zwischen Second- und Third-Party-Punishment im Diktator-Spiel (Fehr & Fischbacher, 2004b, S. 75)**

Die spieltheoretische Voraussage in diesem Zusammenhang ist einfach: Es kommt in keiner der Bedingungen zu einer Bestrafung auf der zweiten Stufe, da keiner der Spieler einen persönlichen Nutzen aus der Bestrafung ziehen kann. Die Spieler A und B der ersten Stufe sehen das voraus und verhalten sich stets unkooperativ, da das die dominante Strategie für sie ist.

Die beobachtete Verteilung stützt die gerechtigkeitspsychologische Hypothese: In beiden Bedingungen wurden Diktatoren, die weniger als 50 Punkte gaben, bestraft (vgl. Abbildung 3). Im Vergleich mit der Bestrafung durch Beteiligte (Second-Party-Punishment) ist die Bestrafung durch Dritte (Third-Party-Punishment) allerdings schwächer, so dass sich unkooperatives Verhalten bei der Bestrafung durch Dritte immer noch lohnte (vgl. Abbildung 4). Durch die Bestrafung werden die Vorteile, die eine unkooperative Person erhält, allerdings deutlich geschmälert. Hier wird klar, welchen Einfluss das Kosten-Bestrafungs-Verhältnis hat. Bei einem günstigeren Verhältnis (z.B. 1:4 statt 1:3) würde sich unfaires Aufteilen auch bei Third-Party-Punishment vermutlich nicht mehr lohnen, da die Bestrafung tendenziell zunimmt, wenn sie billiger wird (vgl. weiter unten den Abschnitt über Kosten der Bestrafung).

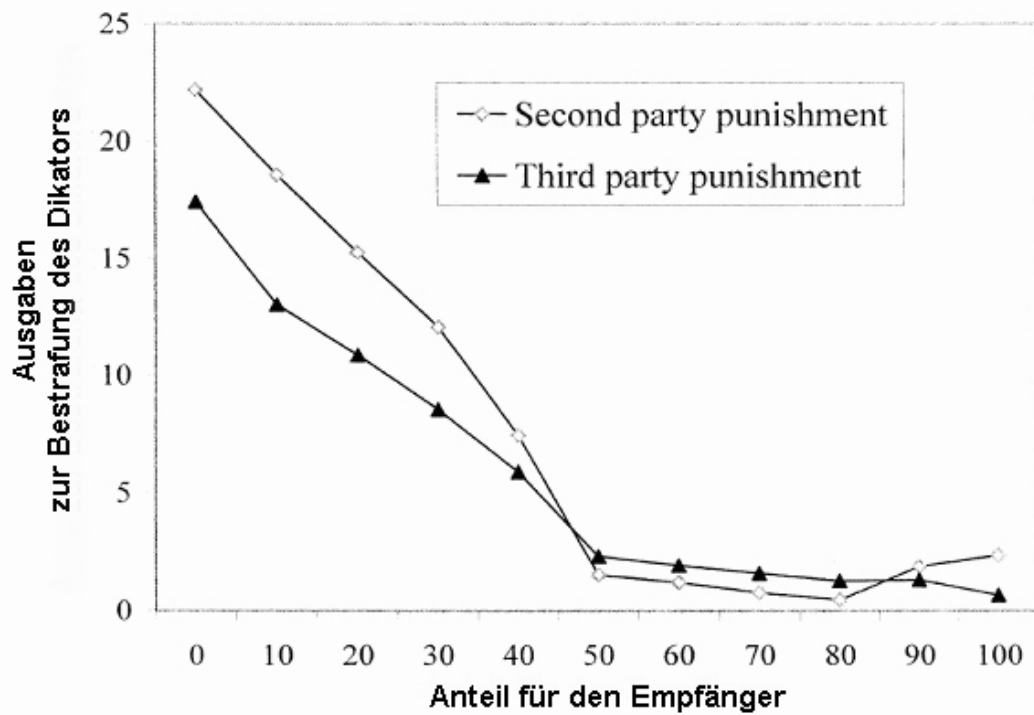


Abbildung 3: Ausgaben für die Bestrafung in Abhängigkeit vom Anteil für den Empfänger (Fehr & Fischbacher, 2004b, S. 78, eigene Übersetzung der Achsenbeschriftung)

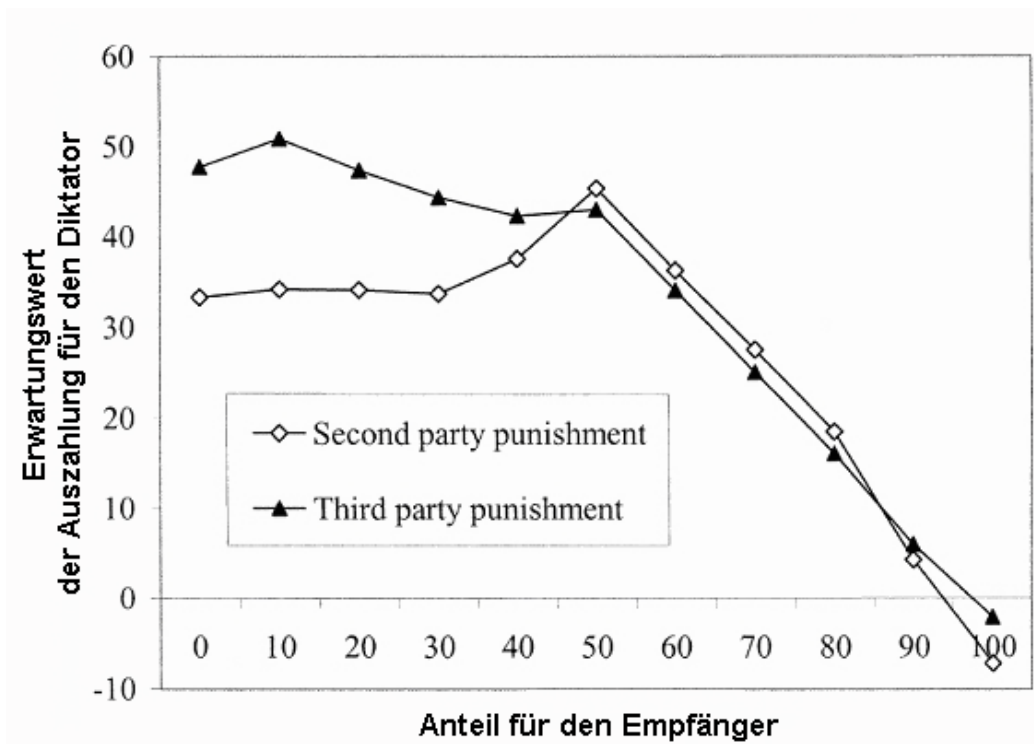
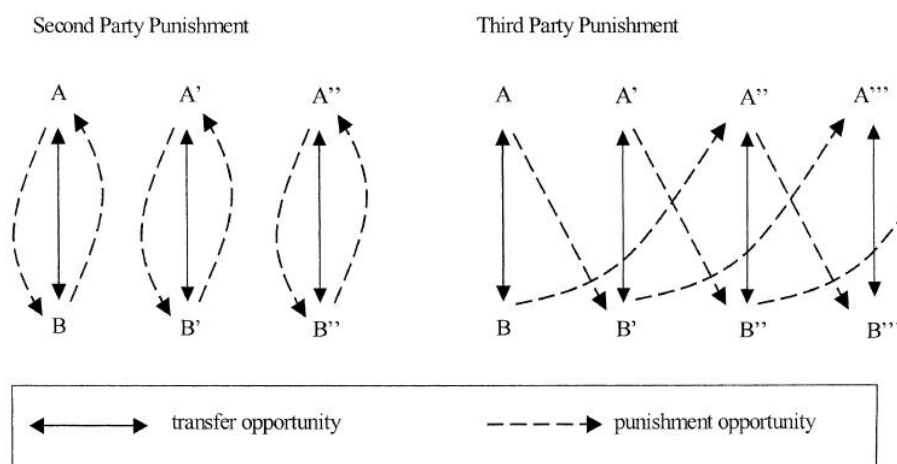


Abbildung 4: Erwarteter Gewinn des Diktators in Abhängigkeit vom Anteil für den Empfänger (Fehr & Fischbacher, 2004b, S. 79, eigene Übersetzung der Achsenbeschriftung)

Für den Vergleich der Bestrafungsarten beim Bruch einer Kooperationsnorm spielten die Versuchspersonen auf der ersten Stufe jeweils zu zweit ein Gefangenens-Dilemma. Auf der zweiten Stufe erhielten beide Spieler zusätzlich zu der Ausstattung aus der ersten Stufe und ihren Gewinnen und 20 Punkte, die sie zur Bestrafung einsetzen konnten. In der Second-Party-Punishment-Bedingung konnten sich die beiden Spieler gegenseitig bestrafen. In der Third-Party-Punishment-Bedingung konnten sie jeweils einen Spieler aus einer anderen Gruppe bestrafen. Es wurde allerdings darauf geachtet, dass zwei Spieler einer bestimmten Gruppe nie Spieler der gleichen (anderen) Gruppe bestrafen konnten (vgl. Abbildung 5).



**Abbildung 5: Aufbau der Bestrafung beim Vergleich zwischen Second- und Third-Party-Punishment im Gefangenens-Dilemma (Fehr & Fischbacher, 2004b, S. 81)**

In beiden Spielen wurde bestraft, und unkooperative Spieler waren stärker betroffen (die Bestrafung kooperativer Spieler kam zwar vor, ist aber vernachlässigbar). Wie schon beim Bruch der Verteilungsnorm zeigte sich allerdings auch hier, dass Second-Party-Punishment stärker war als Third-Party-Punishment. Auch hier lohnte sich unkooperatives Verhalten lediglich in der Third-Party-Punishment-Bedingung. In der Second-Party-Punishment-Bedingung wurden die Gewinne aus unkooperativem Verhalten durch die Bestrafung nicht nur

vollständig aufgezehrt, die unkooperativen Spieler mussten sich im Vergleich zu den kooperativen sogar mit weniger zufrieden geben.

Zusammenfassend lässt sich sagen: Von einem Normbruch selbst betroffene Versuchspersonen bestrafen öfter und stärker als unbetroffene Beobachter (Fehr & Fischbacher, 2004b).

## **4.5. Einflüsse auf Art und Höhe der Bestrafung**

### **4.5.1. Kosten der Bestrafung**

In verschiedenen Experimenten wurden unterschiedliche Kosten der Bestrafung eingesetzt. Es zeigt sich ein negativer Zusammenhang zwischen den Kosten der Bestrafung und der Häufigkeit und Heftigkeit des Einsatzes. Menschen reagieren also durchaus sensibel auf Veränderungen der Höhe und scheinen Kosten-Nutzen-Analysen in diesem Bereich anzustellen (vgl. z. B. Fehr & Fischbacher, 2004a). Rabin (1993, S. 1282) nimmt diesen Punkt explizit als Grundannahme („stylized fact“) in sein Modell auf.

Falk, Fehr und Fischbacher (2005) führten mehrere Gefangenen-Dilemma-Spiele mit drei Spielern und unterschiedlichen Kosten (hohe Kosten vs. niedrige Kosten) der Bestrafung durch. Sie berichten, dass die Bestrafung von unkooperativen Spielern durch kooperative von den unterschiedlichen Kosten kaum beeinflusst wird. Die bösartige Bestrafung von kooperativen Spielern durch unkooperative und die Bestrafung von unkooperativen durch unkooperative verschwindet bei hohen Kosten allerdings vollständig.

Bereits Baker (1974) zeigte in einem Drei-Personen-Experiment, dass die Bereitschaft zur Wiederherstellung von Gleichheit zwischen zwei anderen durch eine dritte Person mit Hilfe von Belohnungen negativ von den Kosten für die Belohnungen abhängt.

Einen Überblick über verschiedene Untersuchungen zu Kosten der Bestrafung liefern Zwick und Chen (1999). Sie kommen zu dem Schluss, dass mit steigenden Kosten weniger bestraft wird und Fair-

ness also einen Preis hat. Allerdings darf bei dieser Betrachtung die starke Wirkung von Emotionen wie Ärger, die mit unfairer Behandlung einhergehen, nicht vernachlässigt werden. Diese Emotionen dürften geeignet sein, Kosten-Nutzen-Analysen und das Streben nach Profit nachhaltig zu beeinflussen oder sogar zu verhindern.

#### **4.5.2. Höhe der Beträge**

Häufig wird kritisiert, dass die Beträge, die in den Experimenten benutzt werden, zu gering seien. Einige Studien zeigen jedoch, dass, insbesondere bei der Frage der Zurückweisung unfairer Angebote im Ultimatum-Spiel, die Höhe der zu verteilenden Summe kaum Einfluss auf die Ergebnisse hat (Camerer, 2003). Entsprechend berichten Hoffman, McCabe und Smith (1996) von einem Experiment, in dem die Verzehnfachung der Einsätze von 10 \$ auf 100 \$ keinen signifikanten Unterschied in den Ergebnissen erbrachte. Lediglich einen kleinen negativen Zusammenhang zwischen der Höhe der Einsätze und den Angeboten in Ultimatum-Spielen fanden Oosterbeek, Sloof und Van de Kuilen (2004).

#### **4.5.3. Strategiemethode**

Mit „Strategiemethode“ (Selten, 1967) bezeichnet man ein Vorgehen, bei dem die Spieler alle möglichen Aktionen und Reaktionen in einem Spiel angeben müssen. Das Gegenstück dazu ist die „Specific Response Methode“, bei der jeder Spieler nur einen Zug macht, bzw. nur auf den spezifischen Zug seines Mitspielers reagiert. Bei Anwendung der Strategie-Methode muss ein Empfänger im Ultimatum-Spiel dementsprechend für alle möglichen Angebote des Aufteilenden angeben, ob er dieses Angebot annimmt oder ablehnt (siehe Fischbacher et al., 2001, für die Messung konditioneller Kooperation im Öffentliches-Gut-Spiel mit der Strategiemethode). Manchmal wird die Abfrage sogar so erweitert, dass sämtliche Spieler alle Strategien für alle möglichen Rollen innerhalb des Spiels angeben müssen. Eine Variante der Strategiemethode z.B. im Ultimatum-Spiel ist die



Abfrage von Mindestangeboten (MAO: minimal acceptable offer), unterhalb derer alle anderen Angebote abgelehnt werden. Bei dieser Vorgehensweise wird implizit davon ausgegangen, dass das Ablehnungsverhalten monoton ist. Verschiedene Studien zeigten allerdings ein nicht-monotones Ablehnungsmuster. Es wurden sowohl zu geringe als auch zu hohe Angebote ablehnt (Güth, Sutter & Schmidt, 2003; Bahry & Wilson, 2006).

Ein Vorteil der Strategiemethode ist die größere Datenmenge, die eine tiefere statistische Auswertung ermöglicht. Da alle Aktionen und Reaktionen abgefragt und beantwortet werden, können auch sehr seltene Aktions- und Reaktionsmuster in die Analyse einbezogen werden (z.B. Fehr & Fischbacher, 2004b).

Roth (1995) sieht in der Anwendung der Strategiemethode eine Veränderung des Spiels, da die Spieler dann über alle möglichen Aktionen (bzw. Reaktionen) nachdenken müssen und nicht nur über eine konkrete. Darüber hinaus ändert sich der zeitliche Ablauf des Spiels, da alle Entscheidungen simultan getroffen werden und nicht mehr als Reaktion auf eine bestimmte Aktion des Mitspielers. Brandts und Charness (2000) finden keine Unterschiede zwischen den beiden Methoden. Im Gegensatz dazu berichten Oosterbeek, Sloof und Van de Kuilen (2004) in einer neueren Meta-Analyse einen signifikanten Einfluss: Bei Anwendung der Strategiemethode waren die Aufteilenden in Ultimatum-Spielen großzügiger, als wenn konkrete Angebote gemacht und darauf reagiert wurde. Außerdem waren die Ablehnungsquoten bei Anwendung der Strategiemethode höher. Falk, Fehr und Fischbacher (2005) finden im Gefangenen-Dilemma-Spiel mit Bestrafung zwar qualitativ die gleichen Bestrafungsmuster bei der Strategiemethode und der Specific Response Methode, die Härte der Bestrafung ist allerdings bei der Reaktion auf ein einzelnes Angebot höher. Fehr und Fischbacher (2004b) verweisen darauf, dass der Einfluss der Emotionen bei der Anwendung der Strategiemethode abgeschwächt sein könnte, da ein Spieler nicht nur auf ein konkretes

Angebot reagieren muss, sondern auf sämtliche möglichen Angebote. Die Frage des Einflusses der Emotionen sehen Fehr und Fischbacher (2004b) als offene Frage an. Welche weiteren Folgen die Strategiemethode auf das Spielverhalten der Versuchspersonen hat, ist nicht abschließend geklärt und weitere Forschung in diesem Bereich ist wünschenswert.

Die Anwendung der Strategiemethode bedeutet nicht, dass die monetären Anreize aufgehoben werden. Die Spieler werden den Rollen zugelost, und dann werden die entsprechenden Aktionen, die sie in ihren Strategien angegeben haben, mit den üblichen monetären Anreizen ausgespielt. Folgendes Beispiel soll das verdeutlichen: Ein Spieler hat angegeben, dass er in der Rolle des Aufteilenden im Ultimatumspiel von 10 Euro 7 für sich behalten und seinem Empfänger 3 Euro anbieten würde. Ein anderer Spieler hat angegeben, dass er in der Position des Empfängers Angebote oberhalb von 2 Euro akzeptieren würde. Werden diese beiden Spieler nun per Los in diesen Rollen zusammen gebracht, erhält der Aufteilende die verlangten 7 Euro und der Empfänger erhält 3 Euro. Hätte der Aufteilende nur 2 Euro angeboten, wären beide entsprechend den Regeln im Ultimatum-Spiel leer ausgegangen.

#### **4.5.4. Anonymität**

Mit steigender Anonymität geht meist geringere Bestrafung einher. Diese Tatsache wird (auch in der Psychologie) häufig so gedeutet, dass die Versuchspersonen fair und gerecht erscheinen möchten, ohne die Kosten dafür tragen zu müssen (für "moral hypocrisy" siehe Batson, Thompson, Seufferling, Whitney & Strongman, 1999). Diese Deutung ist zulässig und häufig sicherlich auch zutreffend. Es kann allerdings auch andere Gründe geben, die für eine geringere Bestrafung bei Anonymität bzw. Ausschluss der Öffentlichkeit, sprechen. Miller und Vidmar (1981) unterscheiden zwei Adressaten der Bestrafung, den Normbrecher und die Mitmenschen. Personen, die mit ihrer Strafe (auch) ihre Anerkennung einer bestimmten Norm

bzw. die Zugehörigkeit zu einer Gemeinschaft, die diese Norm teilt, zum Ausdruck bringen oder ihre Ablehnung der Tat signalisieren möchten, werden unter Anonymität weniger bestraft, da diese Funktion der Bestrafung nicht erfüllt wird. Beim einfachen Vergleich der Bestrafung bei Anonymität und in einer öffentlichen Situation wird vernachlässigt, dass die Bestrafenden mehrere Ziele verfolgen können. Vidmar (2001) weist darauf hin, dass Rache bzw. Strafe auch dazu dienen kann, die Macht und den moralischen Status des Opfers wieder herzustellen und dass dies am besten gelingt, wenn der Täter weiß, warum und von wem er die Strafe empfangen hat.

#### **4.5.5. Intentionen (Spielpartner Mensch vs. Computer)**

Damit eine Handlung oder ein Ergebnis als ungerecht wahrgenommen wird, reicht es nicht aus, dass gegen berechnigte Ansprüche verstoßen wird. Zusätzlich muss die Ursache für den ungerechten Ausgang im Handelnden liegen, und ihm muss Kontrolle über die Handlung sowie eine Intention unterstellt werden können (Mikula, 1993). Allgemein ist davon auszugehen, dass die Bereitschaft zur Bestrafung höher ist, wenn dem Spielpartner eine unfaire Intention unterstellt werden kann (Gollwitzer, 2004; Falk et al., 2003; Turillo et al., 2002). Eine mögliche Erklärung für das Verhalten ist der Einfluss von negativen Emotionen, die entstehen, wenn ein ungünstiges Ergebnis als Resultat einer Intention gesehen werden kann (Heider, 1958).

Intentionen schreibt man nur einem Menschen zu, nicht aber einem Computer. Verschiedentlich wurde untersucht, wie Menschen reagieren, wenn sie mit einem Computer spielen. Es zeigte sich, dass bei einem Computergegner „ungerechte“ Angebote eher akzeptiert werden (z.B. Sanfey et al., 2003; Knoch, Pascual-Leone, Meyer, Treyer & Fehr, 2006; van't Wout et al., 2006).

Sanfey, Rilling, Aronson, Nystrom und Cohen (2003) konnten mit Hilfe von fMRI zeigen, dass es bei unfairen Angeboten durch einen menschlichen Mitspieler, dem Intentionen unterstellt werden, zu

einer stärkeren Aktivierung von Hirnarealen kam, die im Allgemeinen mit negativen Emotionen assoziiert sind. Wurden die ungleichen Angebote von einem Computer gemacht, fiel die Aktivierung viel geringer aus. Darüber hinaus fanden die Autoren einen deutlichen Zusammenhang zwischen der Aktivierung dieser Hirnregionen und dem nachfolgenden Ablehnungsverhalten. Eine hohe Aktivierung ging mit vermehrter Ablehnung unfairer Angebote einher.

#### **4.5.6. Gruppenzugehörigkeit und Gruppengröße**

Wie bereits weiter oben erwähnt, spielt die Frage der Anonymität zwischen den Versuchspersonen eine erhebliche Rolle. Eine Form, die Anonymität zwischen den Versuchspersonen zu verringern, stellen Informationen über die Gruppen- bzw. Familien- oder Staatszugehörigkeit der Mitspieler dar. Bereits minimale Hinweise auf die Gruppenzugehörigkeit können das Verhalten von Personen verändern (Tajfel, Billig, Bundy & Flament, 1971). Verschiedentlich konnte gezeigt werden, dass es je nach Gruppenzugehörigkeit des Normbrechers, des Opfers und des Bestrafenden zu unterschiedlich starker Bestrafung kommt (Bernhard, Fischbacher & Fehr, 2006; Shinada, Yamagishi & Ohmura, 2004).

Bernhard, Fehr und Fischbacher (2006) sowie Goette, Huffman und Meier (2006) zeigen, dass die Kooperationsbereitschaft innerhalb einer Gruppe höher ist, und finden Hinweise darauf, dass Normen in der eigenen sozialen Gruppe stärker durchgesetzt werden. Gintis (2000) weist darauf hin, dass die prosoziale Wirkung starker Reziprozität in kohärenten Gruppen mit einer gemeinsamen Interaktionszukunft am stärksten ist. Hiermit kann vermutlich auch die unterschiedlich starke Bestrafung von Gruppenmitgliedern und Externen erklärt werden. Der positive Einfluss der Gruppenkohäsion auf die Bestrafung von Normbrechern wird von Horne (2001) experimentell nachgewiesen.

#### 4.5.7. Kommunikation

Kommunikation zwischen den Spielern vor oder während des Spiels hat einen großen Einfluss auf das Verhalten. Sie dient z.B. der Koordination des eigenen Beitrags mit dem Beitrag des Spielpartners und dessen Signalisierung (Ostrom, 2000; vgl. auch Wiswede, 1998). Dieser Einfluss von Kommunikation ist insbesondere vor dem Hintergrund der Spieltheorie interessant. Dort wird jede Form von verbaler Äußerung, die keine tatsächliche Festlegung auf eine Strategie ist, als „cheap talk“ – billiges Geschwätz – bezeichnet. Nach dieser Ansicht fühlen sich die „billigen Schwätzer“ nicht an ihr Geschwätz gebunden und da die Mitspieler dies wissen, berücksichtigen sie es auch nicht. Verschiedene empirische Untersuchungen zeigen jedoch, dass sich Menschen durchaus an ihre Aussagen halten und dass andere dies auch antizipieren (z.B. Frank et al., 1993). Kravitz und Guntto (1992) finden, dass die Akzeptanz ungleicher Angebote vom Inhalt mitgelieferter Nachrichten abhing. Wurde mit dem Angebot eine Nachricht übermittelt, welche die Bedürftigkeit des Aufteilenden thematisierte, wurden diese Angebote eher akzeptiert. Wurde allerdings in dem Kommentar die Macht des Aufteilenden betont, sank die Akzeptanzrate deutlich.

Frank, Gilovich und Regan (1993) konnten sogar zeigen, dass ein Zusammentreffen von 30 Minuten vor dem Experiment dazu führte, dass die Versuchspersonen in der Lage waren, die Handlung (Kooperation / Nicht-Kooperation) ihrer zukünftigen Spielpartner überzufällig vorherzusagen, außerdem stieg die Kooperationsrate mit zunehmender Möglichkeit zur Kommunikation vor dem Experiment.

Xiao und Houser (2005) berichten einen interessanten Einfluss der Kommunikation auf das Bestrafungsverhalten bei Second-Party-Punishment. Sie räumten einem Teil der Entscheider in einem Ultimatum-Spiel die Möglichkeit ein, dem Aufteiler eine Nachricht zu schicken und ihm so ihre Emotionen mitzuteilen bzw. ihn auf diesem Weg verbal zu bestrafen. Bestand zusätzlich zur materiellen Be-

strafung unter Kosten diese „billige“ Möglichkeit der verbalen Bestrafung, wurde kaum noch von der teuren Form der Bestrafung Gebrauch gemacht.

#### **4.6. Bösertige Bestrafung**

In verschiedenen Untersuchungen wurde immer wieder ein gewisser Teil an bösertiger Bestrafung (spiteful punishment) festgestellt (vgl. z.B. Falk et al., 2005; Fehr & Fischbacher, 2004b). Darunter verstehen die Autoren die Bestrafung kooperativer Spieler. Bösertige Bestrafung kann sowohl von kooperativen als auch von unkooperativen Mitspielern ausgeführt werden. Negative Emotionen können hier als Erklärung relativ sicher ausgeschlossen werden. Ein solches Verhalten lässt sich allerdings damit erklären, dass mit der üblichen Ausgestaltung der Bestrafung im Verhältnis 1:3 zwischen Einsatz und Bestrafung Unterschiede im Einkommen zwischen Individuen oder Gruppen gezielt erhöht werden können (Shinada et al., 2004; Falk et al., 2005). Wenn diese Möglichkeit nicht besteht (entweder Bestrafung im Verhältnis 1:1 oder nicht-materiell), dann verschwindet die bösertige „Bestrafung“ kooperativer Individuen nahezu vollständig, und es werden fast ausschließlich unkooperative Individuen bestraft (Falk et al., 2005).

Miller und Vidmar (1981) verweisen zudem darauf, dass Bestrafung verschiedene psychologische Funktionen erfüllen und unter anderem auch als Ventil für unterdrückten Sadismus dienen kann.

#### **4.7. Kritik an experimentellen Spielen zur Messung altruistischer Bestrafung**

Die Verwendung experimenteller Spiele in der sozialwissenschaftlichen Forschung hat Gegner und Befürworter. Im folgenden Abschnitt soll der Kritik an dieser Methode Rechnung getragen werden.

#### **4.7.1. Situation künstlich, kontextfreie Darstellung**

Unbestritten ist, dass die kargen spieltheoretischen Situationen relativ künstlich sind und durch die Probanden interpretiert werden müssen. Das kann allerdings genauso als Vorteil des Messinstruments gesehen werden. Häufig sind in den experimentellen Spielen unterschiedliche, teils gegenläufige Verhaltensnormen angesprochen. Wie die Situation von den Versuchspersonen interpretiert wird, hängt dann auch von ihrer Persönlichkeit ab. Es können starke und schwache Situationen unterschieden werden (Mischel, 1976). Vorhersagen auf Basis der situativen Variablen sind am ehesten dann möglich, wenn es sich um starke Situationen handelt, die definitionsgemäß (Mischel, 1977):

- von allen Menschen gleichermaßen wahrgenommen und kodiert werden,
- uniforme Erwartungen über die beste Antwort induzieren,
- ausreichende Anreize für das Ausführen dieser Antwort bieten und
- Anforderungen an die Fähigkeiten stellen, die alle Menschen in gleichem oder zumindest ähnlichem Maß besitzen.

Er nennt als Beispiel eine rote Ampel. Hier gibt es nur eine richtige Reaktion, die Situation ist gut strukturiert und die Menschen, die am Straßenverkehr teilnehmen, verfügen im Allgemeinen über die nötigen Fähigkeiten, adäquat auf diese Situation zu reagieren.

Das gilt offensichtlich nicht für experimentelle Spiele, denn es lässt sich eine große Varianz im Verhalten erkennen, die sich nicht durch die situativen Merkmale erklären lässt. Vielmehr handelt es sich bei ihnen um schwache Situationen, die weniger stark strukturiert sind und keine uniformen Reaktionsmuster hervorrufen (vgl. Camerer & Fehr, 2004, in Bezug auf das Diktator-Spiel).

Eine Untersuchung von experimentellen (schwachen) Spielsituationen ist daher besonders interessant, da davon ausgegangen werden

kann, dass Persönlichkeitseigenschaften (z.B. Ungerechtigkeitssensibilität) die Probanden zu unterschiedlichem Verhalten veranlassen.

#### **4.7.2. Geldbeträge meist zu gering**

„If I had a dollar for every time an economist claimed that raising the stakes would drive ultimatum behaviour toward self-interest, I'd have a private jet on standby all day.”

(Camerer, 2003, S. 60)

Meistens werden bei spieltheoretischen Experimenten Summen in Aussicht gestellt, die einer Entschädigung für den Aufwand der Probanden entsprechen (normalerweise können die Versuchspersonen im Ultimatum-Spiel zwischen 5 und 20 Euro verdienen). Es wurden allerdings auch Untersuchungen zum Ultimatum-Spiel in Ländern, in denen das Einkommen deutlich niedriger liegt als in den USA oder in Europa, durchgeführt. Es ging für die Untersuchten dann um ein Vielfaches eines Tages- oder sogar Monatslohnes. Diese höheren Anreize zeitigten aber keine deutlich unterschiedlichen Ergebnisse und es kam nicht zu einer deutlich geringeren Ablehnungsquote, wie von Kritikern vielfach angenommen wird. Einzig die Varianz sank und es kam zu weniger Ausreißern (Camerer, 2003). Die Bezahlung der Versuchspersonen in den hier berichteten Experimenten war verglichen mit anderen Experimenten dieser Art weder besonders hoch noch besonders niedrig und entsprach einer angemessenen Entschädigung für den Aufwand.

#### **4.7.3. Spiele werden nicht verstanden**

*Auszahlungen und Handlungsalternativen werden nicht verstanden*

Natürlich ist es essentiell, dass die Versuchspersonen den Versuchsaufbau verstehen. Häufig werden den Versuchspersonen Verständnisfragen gestellt, ohne deren richtige Beantwortung sie nicht am Experiment teilnehmen können. Teilweise wird hierauf aber auch verzichtet. Dieser Verzicht ist unzweckmäßig, da darauf geachtet werden



muss, dass die Versuchspersonen mit den gleichen Fakten konfrontiert werden. Verstehen die Versuchspersonen die Situationen (Handlungsmöglichkeiten und Konsequenzen für die Auszahlung) richtig, können Verhaltensunterschiede auf Persönlichkeitsunterschiede zurückgeführt werden.

Manche Autoren versuchen den typischen fallenden Verlauf der Beiträge in einem Öffentlichen-Gut-Spiel mit mangelndem Verständnis zu Beginn des Versuchs zu erklären. Unterschiedliche Lerneffekte werden genannt:

1. Versuchspersonen lernen, was die beste Strategie ist, und kooperieren weniger (Houser & Kurzban, 2002).
2. Versuchspersonen lernen, dass andere nicht kooperieren, und sind frustriert (Andreoni, 1995).
3. Versuchspersonen lernen, auf einem bestimmten (niedrigen) Level zu kooperieren (Ostrom, 2000).

Wenn der Rückgang der Kooperation beim ersten Punkt auf einem mangelhaften Verständnis der Spiel-Situation beruht, sollte er sich nach verbesserten Instruktionen oder Ausschluss der unverständigen Versuchspersonen nicht mehr zeigen. Beim zweiten Punkt wird der Rückgang mit enttäuschten Erwartungen über die Kooperationsbereitschaft der anderen erklärt. Das ist kein mangelndes Verständnis der Situation. Genauso verhält es sich beim dritten Punkt, wo Anpassungen der Kooperationsbereitschaft zu Kooperation auf niedrigem Niveau führen. Wenn das Ziel der Untersuchung nicht die Analyse des zunehmenden Verständnisses der Situation ist, sollten Versuchspersonen, die selbst nach eingehender Erklärung die Situation offensichtlich nicht verstanden haben, von der Analyse ausgeschlossen werden.

*Einmaligkeit der Interaktion wird nicht verstanden*

Häufig wird eingewendet, dass die einmalige anonyme Interaktion in Experimenten nicht dem normalen Interaktionsverhalten der Ver-

suchsteilnehmer entspreche. Daher brächten sie ihre Erfahrungen und Angewohnheiten aus dem Alltag mit ins Labor und könnten zwischen den verschiedenen experimentellen Situationen (einmalig vs. wiederholt) nicht unterscheiden (Hoffman et al., 1996). Kooperation, Bestrafung und Altruismus in den Experimenten resultierten dann aus dem (falschen) Verständnis der Situation. Dies scheint aus mehreren Gründen nicht stichhaltig. Menschen haben und hatten durchaus Erfahrung mit einmaligen anonymen Interaktionen in ihrem Alltag. Außerdem deuten Ergebnisse verschiedener Untersuchungen darauf hin, dass die Versuchspersonen durchaus diskriminieren können (für eine Übersicht über diese Kritik und empirische Ergebnisse, die ihr widersprechen, siehe Fehr & Fischbacher, 2002; für die Frage der evolutionären Anpassung an solche Situationen siehe Hagen & Hammerstein, 2006, sowie die Diskussion im Abschnitt zu ultimativen Erklärungen).

## **5. Experimentelle Untersuchung altruistischer Bestrafung**

### **5.1. Ungerechtigkeitssensibilität, negative Emotionen und altruistische Bestrafung**

Die zwei hier durchgeführten Studien knüpfen empirisch an die Ergebnisse und inhaltlichen Überlegungen der Studien von Fetchenhauer und Huang (2004) sowie an die weiter oben zur Ungerechtigkeitssensibilität (US) geschilderten theoretischen Überlegungen und empirischen Ergebnisse an. Das Ziel der Studien ist, altruistische Bestrafung und deren Ursachen näher zu beleuchten. Insbesondere wird der Versuch unternommen, altruistische Bestrafung durch die dritte Person in einem Third-Party-Punishment-Game mit der Persönlichkeitseigenschaft Ungerechtigkeitssensibilität zu erklären und den Wirkungspfad des Einflusses der Persönlichkeit über moralische Emotionen auf das Verhalten nachzuzeichnen.

Folgende Fragen stehen im Mittelpunkt der Studien 1 und 2:

1. Wie reagiert eine von der Aufteilung unbetroffene dritte Person auf unfaire Angebote eines Aufteilenden an einen Empfänger?
2. Können mögliche Bestrafungen durch die dritte Partei mit US und negativen emotionalen Reaktionen auf die Ungerechtigkeit erklärt werden?

Zu diesem Zweck wurde ausschließlich die Reaktion der dritten Person untersucht.

#### **5.1.1. Studie 1**

In der Vergangenheit wurde bereits wiederholt nachgewiesen, dass Menschen bereit sind, Fremde, von deren ungerechten Handlungen sie nicht betroffen sind, für ungerechtes Verhalten zu bestrafen. Sie tun dies selbst dann, wenn ihnen dabei Kosten entstehen. Fetchenhauer und Huang (2004) konnten zudem zeigen, dass die Unge-

rechtigkeitssensibilität einen wichtigen Beitrag zur Erklärung des Verhaltens in verschiedenen experimentellen Spielen liefert. Davon ausgehend wurde ein leicht verändertes (s. u. bei der Beschreibung der Methode) Drei-Personen-Bestrafungs-Spiel durchgeführt.

Bezüglich der salienten Verteilungsnorm werden hier die gleichen Annahmen wie von Fehr und Fischbacher (2004b) getroffen, die sich so auch aus gerechtigkeitspsychologischen Theorien (z.B. der Equity-Theorie von Walster et al., 1978) ableiten lassen. Es wurde erwartet, dass ein Aufteiler den Betrag gleichmäßig zwischen sich und dem Empfänger verteilt (50/50), da die Versuchspersonen scheinbar zufällig den Rollen zugewiesen worden waren und kein anderer Grund für eine andere Aufteilung spricht. Von den dritten Personen wird angenommen, dass sie Abweichungen von dieser Verteilungsnorm bestrafen, indem sie die Verteilung unter eigenen Kosten ablehnen.

Aus den oben geschilderten theoretischen Zusammenhängen und den weiter oben berichteten empirischen Ergebnissen zur US werden folgende Hypothesen für die Versuchspersonen in der Rolle der dritten Person abgeleitet:

1. Versuchspersonen mit hohen Werten auf  $US_{\text{Nutznießer}}$  lehnen die ungerechten Aufteilungsvorschläge öfter ab als Personen mit niedrigen Werten auf dieser Dimension.

Durch das Akzeptieren einer unfairen Verteilung würden sich die Versuchspersonen zu Mittätern machen, die aus der ungerechten Behandlung des Empfängers Profit schlagen. Insbesondere Personen mit hohen Werten auf  $US_{\text{Nutznießer}}$  werden dies nicht akzeptieren und die Entscheidung ablehnen.

2. Versuchspersonen, die stärkere negative Emotionen (Ärger, Empörung) äußern, lehnen einen unfairen Aufteilungsvorschlag eher ab als diejenigen, die nicht so starke negative Emotionen äußern.

Wie oben beschrieben und verschiedentlich empirisch nachgewiesen, sind es negative Emotionen, die eine Bestrafung verursachen. Entsprechend wird hier davon ausgegangen, dass Versuchspersonen, die vor der Bestrafungsentscheidung stärkere negative Emotionen zeigen, einen unfairen Aufteilungsvorschlag auch eher ablehnen, als diejenigen, die keine so starken negativen Emotionen zeigen.

3. Versuchspersonen, die hohe Werte auf  $US_{\text{Nutznießer}}$  aufweisen, berichten stärkere negative Emotionen.

In Verbindung mit wahrgenommener Ungerechtigkeit tritt bei den Versuchspersonen Ärger auf, dessen Ausmaß mit der Ausgeprägtheit ihrer US korreliert. Daher wird geprüft, ob Personen, die hohe Werte auf der Skala  $US_{\text{Nutznießer}}$  aufweisen, vor der Ablehnungsentscheidung tatsächlich stärkere negative Emotionen äußern.

4. Der Einfluss der Ungerechtigkeitssensibilität wird über negative Emotionen vermittelt. Es wird eine vollständige bzw. teilweise Mediation des Einflusses aus Hypothese 1 erwartet.

Hypothese 4 bezieht sich auf den Wirkungspfad. Verschiedentlich wurde in früheren Studien der Einfluss der Persönlichkeit auf die Bestrafung bzw. Ablehnung unfairer Angebote untersucht (Hypothese 1). In anderen Studien wurde der Einfluss von Emotionen auf die Bestrafung untersucht (Hypothese 2). Soweit mir bekannt ist, wurde jedoch bisher nie der Einfluss eines Persönlichkeitsmerkmals auf die emotionalen Reaktionen bis zum Verhalten untersucht. Dies geschieht in der vorliegenden Untersuchung anhand der vierten Hypothese.

### *Methode*

Die Stichprobe bestand aus N=25 Studierenden der Universität Köln. Die 12 Studentinnen und 13 Studenten studierten ein großes Spektrum verschiedener Fachrichtungen. Das Alter der Versuchsteilnehmer lag zwischen 20 und 38 Jahren, der Altersdurchschnitt lag bei 24.7 Jahren. Die Versuchspersonen erhielten für ihre Teilnahme

an dem ungefähr 30 Minuten dauernden Experiment eine Aufwandsentschädigung von 7.50 Euro, zudem bekamen sie natürlich die anfallenden Erträge aus dem experimentellen Spiel.

Nachdem die Probanden im Versuchsraum an einem Tisch Platz genommen hatten, wurden sie gebeten, kurzfristig noch an einer zweiten Studie teilzunehmen. Innerhalb dieser scheinbar unabhängigen „Studie“ wurden verschiedene soziodemografische Daten sowie die Persönlichkeitsvariable Ungerechtigkeitssensibilität abgefragt. Nach Beendigung dieser „Studie“ wurde mit dem eigentlichen Versuch begonnen. Dieses Vorgehen wurde gewählt, um die Einflüsse der Abfrage der Ungerechtigkeitssensibilität auf das spätere Bestrafungsverhalten zu minimieren.

Danach wurde den Versuchspersonen gesagt, dass das Experiment mit zwei weiteren Personen durchgeführt würde. Der Proband und die zwei anderen Personen würden an der Aufteilung eines Geldbetrages beteiligt sein, wobei die Positionen noch nicht feststünden. Alle beteiligten Personen würden sich gegenseitig nicht kennen und auch nach dem Versuch nicht aufeinander treffen<sup>14</sup>.

Das Spiel wurde den Versuchspersonen wie folgt beschrieben: Eine Person (Aufteiler) teilt 10 Euro zwischen sich und dem Empfänger auf (Diktator-Spiel). Hierbei hat der Aufteiler zwei Möglichkeiten:

1. 5 Euro für sich und 5 Euro für die andere Person
2. 9 Euro für sich und 1 Euro für die andere Person

Die dritte Person hat danach die Aufgabe, dem Aufteilungsvorschlag entweder zuzustimmen oder diesen abzulehnen. Im Falle der Zustimmung würde dem Aufteilungsvorschlag entsprechend aufgeteilt und die dritte Person erhielte zusätzlich 5 Euro vom Versuchsleiter (unabhängig von den 10 aufgeteilten Euro). Im Falle einer Ablehnung erhielten alle drei nichts. Die Kosten der Bestrafung / Ablehnung betragen für die dritte Person also konstant 5 Euro. Hierin besteht

---

<sup>14</sup> In Wirklichkeit wurden die beiden anderen Versuchsteilnehmer nur simuliert s.u.

die Variation zum Spiel von (Fehr & Fischbacher, 2004b), in dem die dritte Person einzelne Bestrafungspunkte zuteilen konnte.

Nach der Erläuterung des Spiels wurden den Versuchspersonen verschiedene Verständnisfragen gestellt. Aufgrund des wirklich einfachen Spiels, das verwendet wurde, gab es keine Verständnisprobleme bei den Probanden.

Auf die Verständnisfragen folgte die scheinbare Auslosung der Spielposition. Aufgrund der Tatsache, dass bestimmte Aufteilungen – z.B. sehr ungleiche, um die es hier aber explizit geht – nur in einem Teil aller Aufteilungsentscheidungen vorkommen, wurden hier die Angebote der Aufteilenden manipuliert. Genauer gesagt, existierte weder der Aufteiler noch der Empfänger. Ihr Verhalten wurde – wie bei Roth und Schoumaker (1983) – den Versuchspersonen simuliert. Allen Versuchspersonen wurde die Rolle der dritten Person „zuge-  
lost“, da auf allen Losen die gleiche Spielposition stand. Keine der Versuchspersonen äußerte diesbezüglich Verdacht.

Im Anschluss an die Festlegung der Spielposition erhielten die Versuchspersonen einen verschlossenen Umschlag, in dem sich die Aufteilungsentscheidung des Aufteilenden befand. Alle Probanden wurden mit einer 9:1-Aufteilung konfrontiert. Die Versuchspersonen wurden gebeten, vor ihrer Entscheidung darüber, ob sie die Aufteilung annehmen oder ablehnen würden, zu „schildern, was ihnen in dieser Situation durch den Kopf geht“ (Die Probanden hatten hierfür ca. eine  $\frac{3}{4}$ -DIN A4-Seite Platz, der aber nie vollständig genutzt wurde). Sie wurden darüber informiert, dass die beiden anderen Personen diese Schilderung mit deren Auszahlungen erhalten würden.

Dann wurden die Probanden gebeten, ihre Entscheidung zu treffen und anzugeben, ob sie die Aufteilung akzeptieren oder ablehnen. Die Aufklärung der Probanden erfolgte nach Abschluss der Untersuchung.

### *Ergebnisse*

Von den 25 Versuchspersonen lehnten 10 die Aufteilung ab. Sie bestraften somit den Aufteilenden und nahmen Kosten in Höhe von 5 Euro auf sich. Es wurde kein Unterschied im (Bestrafungs-)Verhalten von Männern und Frauen gefunden.

Wie oben geschildert, wurden die Versuchspersonen gebeten aufzuschreiben, was ihnen durch den Kopf geht, nachdem sie in der Position des Entscheiders erfahren haben, dass der Aufteilende 9 Euro für sich behält und lediglich 1 Euro an den Empfänger abgibt. Die Versuchspersonen nutzten die Möglichkeit, einen Kommentar abzugeben, recht intensiv, alle Versuchspersonen schrieben einen Kommentar. Die Länge der einzelnen Kommentare variierte beträchtlich (Anzahl Buchstaben:  $M = 286$ ,  $SD = 146.4$ ). Diese Schilderungen wurden von mir und 4 weiteren Ratern bezüglich des dort geäußerten Ärgers und der geäußerten Empörung bewertet.<sup>15</sup> Die Interrater-Reliabilität war für beide Emotionseinschätzungen zufrieden stellend hoch ( $\alpha$ -Ärger = .850;  $\alpha$ -Empörung = .811). Die Rater stimmten also in der Beurteilung der Kommentare gut überein. Weiter oben wurde auf die Unterscheidung zwischen Ärger und Empörung verwiesen. In dieser und der nächsten Studie wird dieser Unterschied bewusst aufgehoben. Zwischen der Einschätzung der Empörung und des Ärgers zeigte sich ein sehr starker Zusammenhang ( $r = .930$ ;  $p < .01$ ). Offensichtlich wurden die Kommentare von den Ratern hinsichtlich dieser beiden Emotionen nahezu gleich eingeschätzt. Daher werden die Fragen nach dem Ärger und der Empörung der Versuchspersonen entsprechend als Zwei-Item-Skala zur Messung negativer Emotionen gesehen. Wenn also im weiteren Verlauf der Studie von „negativen Emotionen“ die Rede ist, ist dieser Wert gemeint. Wie schon beim Verhalten, zeigte sich auch bei den negativen Emotionen kein Unterschied zwischen Männern und Frauen in der Stichprobe.

---

<sup>15</sup> Die Länge der Kommentare wies einen Zusammenhang mit der negativen Emotion auf ( $r = .34$ ;  $p < .05$ ), mit den weiteren untersuchten Variablen zeigte sich kein Zusammenhang.



Gemäß verschiedener früherer Untersuchungen (Fetchenhauer & Huang, 2004; Gollwitzer et al., 2005) wurde angenommen, dass Probanden, die hohe Werte auf  $US_{\text{Nutznießer}}$  aufweisen, die Aufteilung eher ablehnen würden als Personen mit niedrigen Werten auf dieser Dimension. Der erwartete Zusammenhang zeigte sich in den Daten: Personen mit hohen Werten auf  $US_{\text{Nutznießer}}$  lehnten die Verteilung eher ab ( $r = .42$ ;  $p < .05$ ) als Personen, die niedrigere Werte auf dieser Dimension aufwiesen. Dieser Wert entspricht einem mittleren bis großen Effekt nach der Einteilung von Cohen (1988). Cohen schlägt für die Effektstärke von  $r$ -Werten folgende Konvention vor:  $r = .10$  (kleiner Effekt),  $r = .30$  (mittlerer Effekt),  $r = .50$  (großer Effekt). Dieses Ergebnis ist im Einklang mit den Ergebnissen von Fetchenhauer und Huang (2004).

Waren negative Emotionen der Auslöser für die Bestrafung? Ja, die Daten stützen diese Vermutung: Diejenigen Personen, die in den Kommentaren signifikant negativere Emotionen zeigten, lehnten die Aufteilung häufiger ab ( $r = .53$ ;  $p < .01$ ; großer Effekt nach Cohen s. o.).

Des Weiteren wurde, gemäß den oben beschriebenen theoretischen Überlegungen zur Ungerechtigkeitssensibilität, ein Zusammenhang zwischen der Ungerechtigkeitssensibilität und den geäußerten negativen Emotionen (Ärger und Empörung) der Versuchspersonen und der Bestrafung vermutet. Auch dieser Zusammenhang ließ sich empirisch belegen. Versuchspersonen mit hohen Werten auf  $US_{\text{Nutznießer}}$  äußerten stärkere negative Emotionen ( $r = .43$ ;  $p < .05$ ; mittlerer bis großer Effekt) als Personen mit niedrigen Werten auf dieser Dimension.

Nachdem nachgewiesen wurde, dass ein bivariater Zusammenhang zwischen  $US_{\text{Nutznießer}}$  und dem Verhalten, den negativen Emotionen und dem Verhalten sowie  $US_{\text{Nutznießer}}$  und den negativen Emotionen besteht, soll in einem weiteren Schritt versucht werden, die Struktur der Wirkungszusammenhänge genauer zu untersuchen. Es wird an-

genommen, dass der Einfluss von  $US_{\text{Nutznie\ss}er}$  über die negative Emotion auf die Entscheidung wirkt.

Diese Zusammenhangshypothese, die anhand der Abbildung weiter unten nachvollzogen werden kann, ist eine einfache Mediationshypothese (Preacher & Hayes, 2004). Statistisch wird diese Annahme mit Hilfe verschiedener Regressionsrechnungen getestet. Zunächst soll allerdings darauf hingewiesen werden, dass die Anwendung der linearen Regression in diesem Fall umstritten ist. Die Variable „Entscheidung“ ist binär, und daher müsste die Analyse eigentlich anhand einer logistischen Regression erfolgen. Hier wird der Argumentation von Fetchenhauer und Huang (2004) gefolgt, die sich auf Cox und Snell (1970) berufen. Nach dieser Ansicht kann eine lineare Regression dennoch durchgeführt werden, wenn die Verteilung in der abhängigen Variable nicht schief ist als 90:10. Da die Verteilung hier deutlich ausgeglichener ist (10 / 25), wird die lineare Regression berechnet, die Ergebnisse werden allerdings vorsichtig interpretiert.

Die Grundvoraussetzung für einen Mediationseffekt und die nötigen Informationen erhält man durch folgende drei Regressionen<sup>16</sup> (Baron & Kenny, 1986):

1. Regression des Mediators (negative Emotion) auf die unabhängige Variable (hier:  $US_{\text{Nutznie\ss}er}$ ):

Einfluss von  $US_{\text{Nutznie\ss}er}$  auf negative Emotion:

$$\beta = .436 \text{ (} B = .515; SE = .222; p < .05 \text{)}$$

2. Regression der abhängigen Variable (Entscheidung über die Aufteilung) auf die unabhängige Variable (hier:  $US_{\text{Nutznie\ss}er}$ ):

Einfluss von  $US_{\text{Nutznie\ss}er}$  auf Entscheidung:

$$\beta = .417 \text{ (} B = .242; SE = .110; p < .05 \text{)}$$

3. Regression der abhängigen Variable sowohl auf den Mediator als auch auf die unabhängige Variable:

---

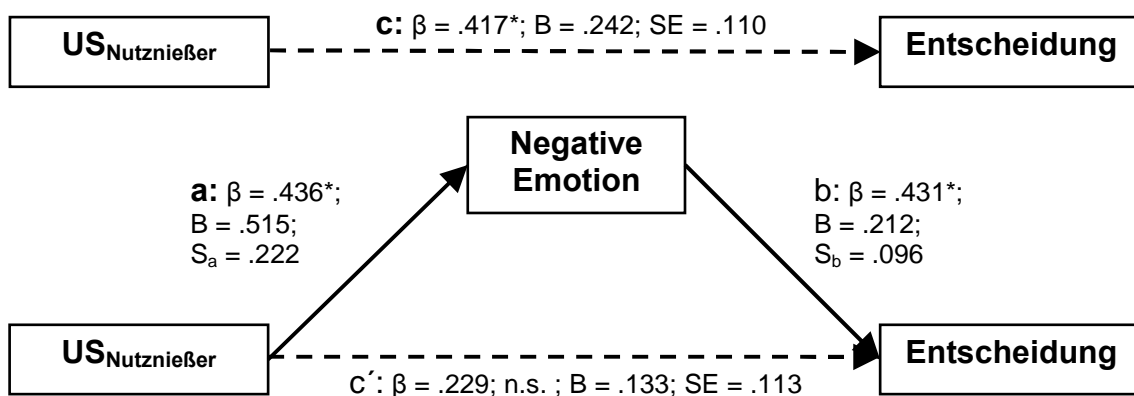
<sup>16</sup> Diese Ergebnisse sind natürlich teilweise identisch mit den bivariaten Zusammenhängen, die weiter oben berichtet werden. Sie erscheinen hier dennoch, da die unstandardisierten B-Koeffizienten sowie die Standardfehler zur Berechnung des Mediationseffektes benötigt werden.

Einfluss von  $US_{\text{Nutznießer}}$  auf Entscheidung:

$$\beta = .229 \text{ (} B = .133; SE = .113; p = .252 \text{)}$$

und Einfluss von „negative Emotion“ auf Entscheidung:

$$\beta = .431 \text{ (} B = .212; SE = .096; p < .05 \text{)}$$



**Abbildung 6: Angenommener Wirkungszusammenhang US - negative Emotionen - Entscheidung mit standardisierten ( $\beta$ ) und unstandardisierten ( $B$ ) Regressionsgewichten und Standardfehlern ( $SE$ ) (\*:  $p < .05$ ; \*\*:  $p < .01$ )**

Folgende Voraussetzungen müssen erfüllt sein:

- In der ersten Gleichung muss es einen Einfluss der unabhängigen Variablen ( $US_{\text{Nutznießer}}$ ) auf den Mediator geben (Voraussetzung erfüllt:  $\beta = .436^*$  s. o.).
- In der zweiten Gleichung muss es einen Einfluss der unabhängigen auf die abhängige Variable geben (Voraussetzung erfüllt:  $\beta = .417^*$ ).
- In der dritten Gleichung muss der Mediator einen Einfluss auf die abhängige Variable haben (Voraussetzung erfüllt:  $\beta = .431^*$ ).

Sind alle diese Voraussetzungen erfüllt, spricht man von einem Mediationseffekt, wenn der Einfluss der unabhängigen Variable in der dritten Gleichung geringer ist als in der zweiten und der Zusammenhang zwischen der unabhängigen Variable und der abhängigen Variable bei Berücksichtigung des Mediators (Gleichung 3) nicht mehr signifikant ist (Baron & Kenny, 1986). Das ist hier der Fall: In

der zweiten Gleichung ist der Einfluss der US ( $\beta = .417^*$ ) höher als in der dritten Gleichung ( $\beta = .229$ ; n. s.) und der Einfluss in der dritten Gleichung ist nicht signifikant. Zusätzlich kann die Signifikanz des Mediationseffektes getestet werden (für verschiedene Tests und genauere Erläuterungen siehe: Preacher & Hayes, 2004). Der Aroian-Test für die vorliegende Mediationshypothese liefert ein nicht signifikantes Ergebnis ( $z = 1.527$ ;  $p = .013$ ). Die Formel für die Berechnung des z-Wertes für den Test lautet (Preacher & Hayes, 2004):

$$z = \frac{a * b}{\sqrt{b^2 * s_a^2 + a^2 * s_b^2 + s_a^2 * s_b^2}}$$

**Formel 1: Berechnung der Signifikanz des Mediationseffektes (a, b: unstandardisierte Regressionskoeffizienten;  $s_a$ ,  $s_b$ : Standardfehler, siehe Abbildung weiter oben)**

Dennoch kann festgehalten werden, dass sich ein Mediationseffekt andeutet, da der Einfluss von US<sub>Nutznießer</sub> auf die Entscheidung nicht mehr signifikant ist, wenn man gleichzeitig die negative Emotion der Versuchspersonen in der Regression berücksichtigt.

### *Diskussion der Studie 1*

Zunächst soll festgehalten werden, dass ein erheblicher Teil der Versuchspersonen bereit war altruistisch zu bestrafen (40%). Sie taten das, indem sie den ungleichen Verteilungsvorschlag ablehnten. Dafür mussten sie auf 5 Euro verzichten, gingen also erhebliche Kosten ein. Die Ablehnung erfolgte, obwohl die untersuchten Personen nicht selbst von der ungleichen Verteilung betroffen waren und sie die anderen beteiligten Personen nicht kannten und auch nicht kennen lernen würden. Es handelte sich also um Third-Party-Punishment. Diese Ergebnisse sind vergleichbar mit Ergebnissen von Fehr und Fischbacher (2004b). Bei Fehr und Fischbacher war der Anteil der Personen, die als dritte Partei altruistisch bestrafen, sogar noch etwas höher. Sie berichteten, dass ca. 60% der Probanden altruis-

tisch bestrafen, während es in der vorliegenden Untersuchung nur 40% (10/25) waren. Ein Grund dafür, dass in der vorliegenden Untersuchung weniger bestraft wurde, könnte der höhere (konstante) Preis für die Bestrafung sein. Während bei Fehr und Fischbacher eine graduelle Bestrafung möglich war, konnten die Probanden hier nur entweder annehmen oder ablehnen, was bei Ablehnung zu Kosten von 5 Euro führte.

Ebenso wie bei Fetschenhauer und Huang (2004) ist es auch hier wieder gelungen, die Dimension  $US_{\text{Nutznießer}}$  erfolgreich zur Verhaltensvorhersage einzusetzen. Es zeigte sich, dass Probanden mit hohen Werten auf der  $US_{\text{Nutznießer}}$  – Dimension eher bereit waren, die ungerechte Aufteilung abzulehnen und die Aufteilenden zu bestrafen. Die moralische Motivation dieser Versuchspersonen war offensichtlich so stark, dass sie bereit waren, hierfür auf 5 Euro zu verzichten.

Warum bestrafen die Versuchspersonen mit den hohen Werten auf der Nutznießerdimension eher? Diese Versuchspersonen berichteten stärkere negative Emotionen und diejenigen, die stärkere negative Emotionen berichteten, bestrafen auch eher. Insofern sind diese Ergebnisse im Einklang mit den obigen Ausführungen, die negative Emotionen als notwendige Bedingung für Bestrafung bzw. Wiederherstellung von Gerechtigkeit sehen (vgl. Hassebrauck, 1984; Pillutla & Murnighan, 1996; Elster, 1998; Brandstätter & Königstein, 2001; Fehr & Gächter, 2002; Fehr & Fischbacher, 2004b; Skarlicki & Kulik, 2005; Xiao & Houser, 2005; Masclet & Villeval, 2006; van't Wout et al., 2006).

Kritisch muss angemerkt werden, dass die Versuchspersonen zum Zeitpunkt des Kommentars schon wussten, dass sie sich gleich entscheiden mussten und ihre Entscheidung häufig in den Kommentar schrieben. Aus diesem Grund kann es sein, dass sich die Versuchspersonen schon entschieden hatten, als sie den Kommentar schrieben. Es kann also nicht sicher festgestellt werden, dass die Emotion tatsächlich der Entscheidung voran ging. Es kann außerdem nicht

ausgeschlossen werden, dass die Rater bei ihrer Beurteilung der Emotionen die erwähnte Entscheidung mit berücksichtigten, so dass der Zusammenhang zwischen negativer Emotion und Verhalten aus diesem Grund möglicherweise zu hoch sein könnte. Ein weiterer Kritikpunkt betrifft die Erhebung der US. Sie wurde unmittelbar vor dem eigentlichen Experiment – getarnt durch eine Cover-Geschichte – erhoben. Wünschenswert wäre eine Erhebung der US deutlich früher vor dem Experiment.

Die beiden anderen Spieler wurden hier nur simuliert, da eine derartig ungleiche Verteilung zwar vorkommt, allerdings nur relativ selten. Normalerweise wird in einem solchen Fall, in dem es darum geht, schwach besetzte Zellen zu untersuchen, auf die Strategiemethode zurückgegriffen. Wie bereits weiter oben erwähnt, wurde allerdings verschiedentlich die Vermutung geäußert, dass die Strategiemethode zu geringeren emotionalen Reaktionen führt als die direkte Reaktion auf ein tatsächliches Angebot (vgl. z.B. Fehr & Fischbacher, 2004b). In der vorliegenden Untersuchung stehen die Emotionen aber im Mittelpunkt des Interesses. Daher ist es entscheidend, von den Probanden die Reaktion auf ein scheinbar reales, besonders unfaires Angebot zu erheben. Wenn aus dem genannten Grund die Strategiemethode nicht angewendet werden kann, würde hier ohne Simulation der anderen beiden Spieler ein unverhältnismäßig hoher Aufwand entstehen. Alle Dreier-Gruppen, in denen der Aufteilende 50/50 aufteilt, sind für die Analyse vollständig wertlos und aufgrund der Bestrafungsmöglichkeit ist von einem relativ hohen Anteil an 50/50-Verteilungen auszugehen. Ein weiterer Vorteil der hier angewendeten Vorgehensweise besteht darin, dass zwangsläufig zwischen dem Aufteilenden, dem Empfänger und der dritten Person völlige Anonymität herrscht. Diese drei Personen werden sich – wie in den Instruktionen angegeben – garantiert nie treffen und kennen lernen.

Zusammenfassung der Ergebnisse der Studie 1: Es konnte gezeigt werden, dass Versuchspersonen mit hohen Werten auf der Dimen-

sion US<sub>Nutznießer</sub> zu stärkeren negativen emotionalen Reaktionen auf Ungerechtigkeit neigen als Personen mit niedrigeren Werten auf dieser Dimension. Diese negativen Emotionen veranlassen die Versuchspersonen, die ungerechte Aufteilung unter eigenen Kosten abzulehnen und die Aufteiler zu bestrafen.

### **5.1.2. Studie 2**

In Studie 1 wurde ein Mediationsmodell entwickelt und geprüft. Studie 2 baut auf diesen Ergebnissen auf. Es soll also zunächst untersucht werden, ob sich der gleiche Zusammenhang US<sub>Nutznießer</sub> – negative Emotion – Bestrafungsentscheidung auch mit einem leicht veränderten Untersuchungsdesign nachweisen lässt. Für diese Fragen wurden die aus Studie 1 bekannten Hypothesen erneut geprüft. Darüber hinaus sollten selbstverständlich die Kritikpunkte an Studie 1 ausgeräumt und weiterführende Analysen vorgenommen werden. Insbesondere steht die Motivation derer, die nicht bestrafen, im Zentrum des Interesses. Sind diese Menschen nur eigennutzorientiert und bestrafen nicht, weil sie die Kosten scheuen oder gibt es auch andere Gründe?

Um detaillierte Aussagen über das Verhalten der verschiedenen Gruppen (Bestrafende / Nicht-Bestrafende) treffen zu können, wurde die Stichprobe stark vergrößert.

Weil in Studie 1 die Ungerechtigkeitssensibilität unmittelbar vor dem eigentlichen Experiment erhoben wurde, könnte es sein, dass die Abfrage zu einer Voraktivierung gerechtigkeitsrelevanter Kognitionen und Emotionen geführt hat, die das Verhalten beeinflussten. In Studie 2 erfolgte diese Messung daher mehrere Wochen (mind. 4 Wochen) vor dem Experiment in anderem Zusammenhang, so dass eine Beeinflussung sicher ausgeschlossen werden kann. Für die Versuchspersonen war es nicht möglich die Messung der US dem Experiment zuzuordnen, da zwischenzeitlich noch andere Persönlichkeitsmaße erfasst und eine andere Untersuchung durchgeführt worden war.

Ein zweiter Kritikpunkt war, dass die Versuchspersonen beim Schreiben des Kommentars bereits wussten, dass sie gleich entscheiden würden, ob sie die Aufteilung annehmen oder ablehnen. Es kann also nicht sicher gesagt werden, ob die Schilderung – und somit die geäußerten Emotionen – tatsächlich der Entscheidung vorangingen. Es spricht zwar einiges dafür, dass die emotionale Reaktion automatisch aktiviert wird und kognitiven Abwägungen vorausgeht (Haidt, 2001), um aber sicher zu sein, wurden Kommentar und Entscheidung in Studie 2 getrennt. In Studie 2 wurden die Versuchspersonen zunächst über die Situation informiert und gebeten zu schildern, was ihnen durch den Kopf geht. Erst danach wurden sie darüber informiert, dass sie Person A bestrafen können.

Eine weitere Veränderung gegenüber dem Design von Studie 1 besteht darin, dass die Versuchspersonen nun die Möglichkeit hatten, graduell zu bestrafen. Es wurde keine reine 0/1-Entscheidung mehr gefordert.

### *Methode*

Die Stichprobe bestand aus N=112 Studierenden der Universität Köln. Aufgrund von Verständnisschwierigkeiten des experimentellen Spiels mussten 19 Probanden von den weiteren Analysen ausgeschlossen werden und zwei weitere, da wichtige Angaben (u. a. US, Alter, Geschlecht) fehlten. Bezüglich des Alters und des Anteils an Männern und Frauen sowie der für die Analysen relevanten Variablen (US, Bestrafungsverhalten, negative Emotion) unterscheiden sich die ausgeschlossenen Probanden nicht von der analysierten Stichprobe. Alle Analysen beziehen sich also auf N=91 Versuchspersonen. Die 58 Studentinnen und 33 Studenten studierten ein großes Spektrum verschiedener Fächer. Das Alter der Versuchsteilnehmer lag zwischen 19 und 42 Jahren, der Altersdurchschnitt betrug 23 Jahre (SD = 3.44). Die Versuchspersonen erhielten für die Teilnahme an den drei Versuchsterminen eine Aufwandsentschädigung von insgesamt 25 Euro. Außerdem erhielten sie natürlich anfallende Erträge aus dem experi-



mentellen Spiel. Das Experiment wurde mit Hilfe eines Computers durchgeführt.

Die Versuchspersonen wurden auf dem Gelände der Universität zu Köln angesprochen und gefragt, ob sie an einer Reihe von Versuchen des Instituts für Wirtschafts- und Sozialpsychologie teilnehmen wollten. Direkt beim ersten Kontakt füllten die Versuchspersonen einen Fragebogen aus, der unter anderem die US-Skala enthielt. Mehrere Wochen später fand dann für die Versuchspersonen beim zweiten Termin ein Experiment statt, das mit dem hier berichteten nicht in Verbindung steht.

Beim dritten Termin wurde die hier referierte Bestrafungsstudie durchgeführt. Die Zusammenführung der Daten erfolgte bei voller Wahrung der Anonymität über ein Codewort (6 Zeichen: erster und zweiter Buchstabe des Vornamens der Mutter, Tag der Geburt, erster und zweiter Buchstabe des Vornamens des Vaters), das die Versuchspersonen selbst erstellten. Sie wurden einzeln in den Versuchsraum geführt und nahmen vor einem Computer Platz. Daneben lag ein Briefumschlag, auf dessen nicht sichtbarer Seite die Anweisung zum Eintragen des persönlichen Codewortes stand. Auf diesen Umschlag wurde nicht gesondert hingewiesen, er wurde im weiteren Verlauf des Versuchs angesprochen. Außerdem wurden den Versuchspersonen fünf einzelne Euromünzen ausgehändigt und es wurde ihnen gesagt, dass sie diese zusätzlich zu ihrer Aufwandsentschädigung erhielten. Danach startete das Experiment am Computer und der Versuchsleiter verließ den Raum. Alle weiteren Instruktionen gingen aus dem programmierten Skript<sup>17</sup> des Experiments hervor.

Nach einer kurzen Begrüßung wurde den Versuchspersonen der Inhalt des Experiments geschildert. Es gehe um die Aufteilung von Geldbeträgen zwischen Personen, die sich nicht kennen und auch nicht kennen lernen werden.

---

<sup>17</sup> Das Experiment wurde mit der Unipark-Software der Firma Globalpark umgesetzt.

Die Versuchspersonen wurden gebeten, ihr Codewort einzugeben. Danach erfolgte die Beschreibung der Situation (Originalinstruktionen):

„Im Moment wird am Institut noch ein anderes Experiment durchgeführt. Dort geht es um die Aufteilung eines Betrages von 10 Euro zwischen zwei Personen, die sich nicht kennen und sich auch während des Experiments nicht kennen lernen.

Eine der beiden Personen (Person A) wird zufällig bestimmt, die Aufteilung der 10 Euro vorzunehmen.

Person A erhält also 10 Euro vom Versuchsleiter und die Aufgabe, diese 10 Euro aufzuteilen.

Hierbei hat Person A zwei vorgegebene Aufteilungsalternativen:

1. Alternative: Person A behält 5 Euro für sich und gibt 5 Euro an Person B.
2. Alternative: Person A behält 10 Euro für sich und gibt 0 Euro an Person B.


Person B kann keinen Einfluss auf die Aufteilung nehmen und bekommt das, was Person A für sie vorsieht.

Beide Personen erhalten die gleiche Beschreibung der Situation und sind so über die Handlungsalternativen der anderen Person informiert.“

Wie in Studie 1 existierten die anderen beiden Personen nur scheinbar. Ihr Verhalten wurde simuliert.

Im Anschluss an die Beschreibung wurden den Versuchspersonen zwei Verständnisfragen zu diesem Diktator-Spiel gestellt. Nur wer diese Fragen korrekt beantwortete, konnte das Experiment fortsetzen.

Nachdem sicher gestellt war, dass die Probanden die Situation verstanden hatten, wurden sie mit dem vermeintlichen Aufteilungsvorschlag von Person A konfrontiert. Dieser sah für alle gleich aus:



4%

### Aufteilungsentscheidung von Person A

Codewort der Person A:

**Person A behält für sich  Euro und gibt  Euro an Person B**

---

Wenn Sie die Aufteilung der Person A zur Kenntnis genommen haben, machen Sie bitte auf der nächsten Seite weiter.

**Abbildung 7: Aufteilungsentscheidung der simulierten Person A (Studie 2)**

Nachdem die Probanden den Aufteilungsvorschlag zur Kenntnis genommen hatten, wurden sie gebeten zu schildern, was ihnen „in dieser Situation durch den Kopf geht“. Die Formulierung war bewusst offen und neutral gehalten, um nicht Gerechtigkeitsurteile oder ähnliches zu provozieren. Die Versuchspersonen hatten im Prinzip unbegrenzt Platz, um ihren Kommentar abzugeben. Wichtig ist, dass zu diesem Zeitpunkt von einer Bestrafungsmöglichkeit oder ähnlichem noch nicht die Rede war (Trennung von Kommentar und Bestrafungsentscheidung).

Erst im Anschluss an den Kommentar wurde den VPn angekündigt, dass sie mit den zusätzlichen 5 Euro, die sie erhalten hatten, in die Auszahlung von Person A eingreifen könnten (Wörter, wie „bestrafen“, „unfair“, „ungleich“ und dergleichen wurden selbstverständlich vermieden). Die VPn wurden ausdrücklich darauf hingewiesen, dass sie die fünf Euro ganz oder teilweise für einen Abzug einsetzen könnten, diese aber auch für sich behalten könnten. Darüber hinaus wurde

erklärt, dass jeder zum Abzug eingesetzte Euro einen Abzug von zwei Euro bei Person A zur Folge habe. Kosten und Abzug standen im Verhältnis 1:2 und waren daher also höher als bei Fehr und Fischbacher (2004b). Die Begründung dafür ist in der einfachen Rechnung zu sehen. Auf diese Weise war für die Versuchspersonen gut nachvollziehbar, dass sie bei voller Bestrafung (5 Euro) dem Aufteilenden alles wegnehmen würden.

Um sicher zu sein, dass die VPn sich auch über die Konsequenzen ihres Handelns im Klaren waren, folgten auf die Beschreibung der Abzugsmöglichkeit zunächst zwei Beispiele mit den dazugehörigen Auszahlungen für alle drei Personen und danach zwei Kontrollfragen. Bei der ersten Kontrollfrage mussten die Auszahlungen für alle drei Personen für eine 10/0-Aufteilung ohne Abzug und für eine 10/0-Aufteilung mit 4 Euro Abzug eingetragen werden. Versuchspersonen, die diese Kontrollfragen nicht richtig beantworteten, wurden von der Analyse ausgeschlossen.

Anschließend wurde der Betrag, der zum Abzug eingesetzt werden sollte, abgefragt. Den VPn wurde gesagt, dass sie einen Abzug in 1-Euro-Schritten vornehmen könnten. Wenn sie nichts abziehen wollten, könnten sie „0“ eintragen. Danach wurden sie kurz an die Kosten und Konsequenzen des Abzugs erinnert, bevor die eigentliche Abfrage erfolgte. Der nächste Bildschirm zeigte ihnen die genauen Konsequenzen, nämlich wie viel Person A abgezogen wurde und wie viel von ihren zusätzlichen 5 Euro noch übrig war.

Die Verdeutlichung der Konsequenzen war wichtig, da im Anschluss daran ein zweiter Kommentar erhoben wurde. Je nach Verhalten (Abzug / kein Abzug) wurden die Probanden nach Gründen gefragt, warum sie nichts abgezogen bzw. etwas abgezogen hatten.

Zuletzt wurde den Versuchspersonen die Frage gestellt, wie sie in der Rolle des Aufteilenden (Position von Person A) die 10 Euro aufgeteilt hätten, wobei der vollständige Alternativenraum zur Verfügung stand (i. e. 10/0 bis 0/10, jeweils in 1-Euro-Schritten). Danach sollten sie

auch diese Entscheidung kurz begründen. Das Experiment endete mit der Abfrage des Geschlechts, des Studiengangs und des Einkommens.

### *Ergebnisse*

Von den 91 VPn bestraften 63 gar nicht, 5 mit 1 Euro, 12 mit 2 Euro, 9 mit 3 Euro, 1 VP bestrafte mit 4 Euro und 1 VPn setzte die ganzen 5 Euro zum Abzug ein. Es wurde kein Unterschied im (Bestrafungs-) Verhalten zwischen Männern und Frauen gefunden. Da in der Stichprobe nur ca. 30% der Probanden bestraften und diese 30% sich auf die verschiedenen Bestrafungslevel verteilten, waren diese Zellen recht dünn besetzt. Da zudem eine Analyse über verschiedene Bestrafungsniveaus keine zusätzlichen Erkenntnisse erbrachte, wird die Variable Bestrafungsentscheidung in zwei Gruppen eingeteilt: Bestrafung vs. keine Bestrafung. Die Auswertung erfolgt dann analog zur Studie 1 mit dieser dichotomen Variable.

In dieser Studie gaben die Versuchspersonen mehrere Kommentare ab. Sie wurden einmal nach ihrer Meinung gefragt, nachdem sie die ungleiche Aufteilung der Person A zur Kenntnis genommen hatten. Beim zweiten Kommentar sollten sie begründen, warum sie einen Abzug vorgenommen bzw. keinen Abzug vorgenommen hatten.

Zunächst sollen nun analoge Analysen zu Studie 1 berichtet werden, um zu sehen, ob sich deren Ergebnisse replizieren lassen. Der erste Kommentar wurde dem Kommentar aus Studie 1 entsprechend ausgewertet. Wieder nutzten die VPn die Möglichkeit der Kommentierung unterschiedlich intensiv. Alle schrieben einen Kommentar, der allerdings sehr stark in der Länge variierte (Anzahl Buchstaben:  $M = 222.9$ ;  $SD = 191.4$ ). Die Kommentare wurden von mir und zwei weiteren Ratern bezüglich des geäußerten Ärgers und der Empörung bewertet. Die Interrater-Reliabilität war für beide Emotionseinschätzungen zufrieden stellend hoch ( $\alpha$ -Ärger = .90;  $\alpha$ -Empörung = .91). Wie bereits in Studie 1, bestand auch hier wieder ein sehr starker Zusammenhang zwischen der Einschätzung der Empörung und der des

Ärgers ( $r = .92$ ;  $p < .01$ ), so dass auch hier wieder der Mittelwert beider Einschätzungen als Wert für die negative Emotion benutzt wurde. Wie in Studie 1 war auch hier kein Unterschied bei den negativen Emotionen zwischen Männern und Frauen in der Stichprobe erkennbar.

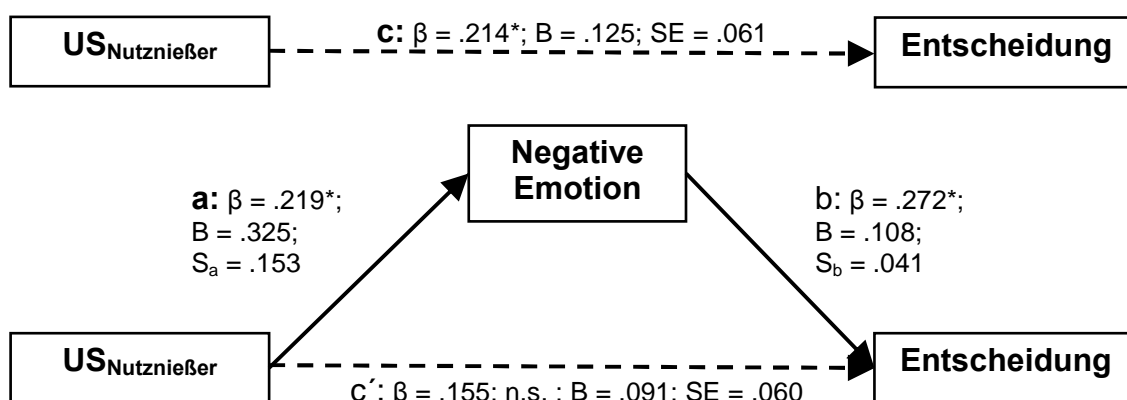
Analog zu Studie 1 wurde wieder ein positiver Einfluss von  $US_{\text{Nutznießer}}$  auf das Bestrafungsverhalten angenommen. Dieser Zusammenhang ließ sich auch in der Stichprobe der Studie 2 nachweisen: Probanden mit höheren Werten auf  $US_{\text{Nutznießer}}$  bestraften die ungleich aufteilenden Personen eher als Probanden mit niedrigen Werten ( $r = .21$ ;  $p < .05$ ; kleiner bis mittlerer Effekt nach Cohen). Auffällig ist, dass der Zusammenhang in dieser Stichprobe deutlich geringer ist als in Studie 1, dort ergab sich ein mittlerer bis großer Effekt. Hierbei muss allerdings berücksichtigt werden, dass die US in Studie 2 mehrere Wochen vor dem Experiment erhoben worden war, in Studie 1 erfolgte die Abfrage der US unmittelbar vor dem Experiment.

In Studie 1 wurde ein mittlerer bis großer Effekt der negativen Emotionen auf die Bestrafung gefunden. Diejenigen Personen, die in ihren Kommentaren stärkere negative Emotionen berichteten, waren auch eher zur Bestrafung bereit. Auch dieses Ergebnis konnte repliziert werden: Es zeigte sich ein Zusammenhang mittlerer Stärke zwischen der negativen Emotion und der Bestrafungsentscheidung ( $r = .31$ ;  $p < .05$ ; mittlerer Effekt nach Cohen). Wie bereits erwähnt, wurden die Kommentare von der Entscheidung zur Bestrafung klar getrennt. Es steht also fest, dass die negativen Emotionen der Bestrafungsentscheidung zeitlich vorausgingen, so dass zur Kausalität klare Aussagen getroffen werden können. Die Hypothese, dass die Bestrafungsentscheidung durch negative Emotionen ausgelöst wird, wird durch die Daten gestützt.

Der in Studie 1 festgestellte Zusammenhang zwischen  $US_{\text{Nutznießer}}$  und den negativen Emotionen in den Kommentaren ließ sich in Studie 2 replizieren. Auch hier wurde ein positiver Zusammenhang zwischen

US<sub>Nutznießer</sub> und den negativen Emotionen gefunden. Probanden, die hohe Werte auf dieser Dimension aufwiesen, äußerten auch in dieser Studie negativere Emotionen, wenn sie mit der ungleichen Aufteilung konfrontiert wurden ( $r = .22$ ;  $p < .05$ ; kleiner bis mittlerer Effekt).

Bei der Prüfung des Mediationsmodells gelten grundsätzlich die gleichen Annahmen und Einschränkungen wie in Studie 1.



**Abbildung 8: Angenommener Wirkungszusammenhang US - negative Emotionen – Entscheidung mit standardisierten ( $\beta$ ) und unstandardisierten ( $B$ ) Regressionsgewichten und Standardfehlern ( $SE$ ) (\*:  $p < .05$ ; \*\*:  $p < .01$ )**

Sämtliche Voraussetzungen für einen Mediationseffekt sind erfüllt, und die Verringerung des Einflusses von US<sub>Nutznießer</sub> von .21 ( $p < .05$ ) auf .16 (n. s.) legt diesen Schluss nahe. Dennoch konnte auch in dieser Studie kein signifikanter Mediationseffekt nachgewiesen werden ( $z = 1.59$ ;  $p = .11$ ). Genau wie in Studie 1 deutet sich allerdings ein hypothesenkonformer Effekt an, der offensichtlich aber recht schwach ist. Dennoch stützen die gefundenen bivariaten Zusammenhänge den angenommenen Wirkungszusammenhang.

Auch das Verhalten derer, die nicht bestraft haben, wurde genauer untersucht. Eine einfache Einteilung in Menschen, die sich für Gerechtigkeit interessieren, und Menschen, die sich nicht für Gerechtigkeit interessieren, läge nahe. Erstere würden ungerechte Menschen bestrafen, letztere nicht. Diese Einteilung greift jedoch zu kurz.

Menschen können durchaus an Gerechtigkeit und an ihren Mitmenschen im Allgemeinen interessiert sein und dennoch eine Bestrafung aus verschiedenen Gründen ablehnen, z.B. weil sie grundsätzlich gegen eine Bestrafung sind oder weil sie in einer reinen Bestrafung, ohne Kompensation für das Opfer, keinen Nutzen sehen (vgl. hierzu auch die Gedanken zur ablehnenden Haltung einiger Menschen gegenüber dem Prinzip des "positive retributivism" von Mackie, 1985, S. 208). In der vorliegenden Studie war klar, dass die drei Personen sich nicht wieder treffen würden. Eine Kompensation für das Opfer (Person B) war nicht vorgesehen und der direkte Einfluss der Bestrafung auf das zukünftige Verhalten der Person A war nicht überprüfbar.

Zur Analyse der Begründungen der Personen, die nicht bestraft hatten, wurden deshalb die Kommentare, die die VPn nach der Entscheidung aufgeschrieben hatten, inhaltsanalytisch ausgewertet. Die Rater hatten die Aufgabe, die Kommentare auf zehn verschiedenen Dimensionen zu bewerten. Die Interrater-Reliabilität war für alle Items zufrieden stellend (alle  $\alpha$ -Werte  $> .70$ ). Eine Faktorenanalyse (Hauptkomponentenanalyse, Varimax-Rotation) über die zehn Items ergab eine 3-Faktorlösung (Eigenwerte: 2.42; 2.3; 1.48; 61.29 % erklärte Varianz), die sich inhaltlich sehr gut interpretieren ließ. Beim Versuch, hierauf aufbauend drei Skalen zu bilden, stellte sich heraus, dass dies nur für zwei Skalen möglich war (aufgrund der mangelnden Reliabilität). Der dritte Faktor geht dennoch über das stärkste Einzelitem in die Analyse ein [in den eckigen Klammern stehen die Items der jeweiligen Skala]:

1. Skala: „Moralische Bedenken“ (3 Items Cronbach's  $\alpha = .80$ ):  
Kommentare, die mit hohen Werten auf dieser Skala bewertet wurden, zeugen von einer Zwickmühle, in der sich die Verfasser dieser Kommentare befunden haben. Sie haben nicht bestraft, sehen aber, dass es hierfür durchaus gute Gründe gibt [...nennt sowohl Argumente für als auch gegen das Bestrafen /



...fände es gerecht in dieser Situation zu bestrafen]. Dieses moralische Dilemma führt bei ihnen zu einem schlechten Gefühl [...fühlt sich unwohl, weil sie nicht bestraft hat].

2. Skala: „Eigennutz“ (3 Items Cronbach's  $\alpha = .59$ ): Kommentare, die auf dieser Dimension hoch bewertet wurden, deuten auf Personen, die zunächst an sich selbst denken [... bestraft nicht, weil sie sich sonst selbst schadet]. Diese Personen versuchen aber, Argumente zu finden, die ihr Eigennutzstreben verschleiern [...bestraft nicht, weil es B nichts bringen würde]<sup>18</sup>. Außerdem verweisen diese Personen darauf, dass alle – sie eingeschlossen – einen Nachteil durch die Strafe erleiden [...bestraft nicht, weil alle dann schlechter dran wären].
3. Einzelitem: „Helfen: ja, Bestrafen: nein“: Die Kommentare der Verfasser, die angaben, nicht an Strafe, jedoch an Kompensation für Person B interessiert zu sein, wurden auf diesem Item mit einem hohen Wert bewertet [...bestraft nicht, hätte aber Geld an Person B transferiert].

Wie hängen diese Dimensionen mit der Persönlichkeit der Versuchspersonen zusammen? Eine erste Hypothese ist, dass Personen, die bestrafen und diejenigen, die nicht bestrafen, danach aber moralische Bedenken äußern, sich von der Persönlichkeitsstruktur her ähneln. Beide haben ein starkes Gerechtigkeitsempfinden und sehen es grundsätzlich als gerecht an, den unfair Aufteilenden zu bestrafen.

Um hier ein umfassenderes Bild der Persönlichkeit zeichnen zu können, wird bei den folgenden Analysen auch die Dimension  $US_{\text{Opfer}}$  betrachtet. Frühere Untersuchungen zeigten bereits, dass Personen mit hohen Werten auf  $US_{\text{Opfer}}$  eher versuchen, Gerechtigkeit gegenüber sich selbst durchzusetzen und dabei auch von eigennützigen Beweggründen getrieben sind (siehe oben im Abschnitt zur  $US_{\text{Fetthauer}}$  & Huang, 2004; Gollwitzer et al., 2005; Schmitt et al.,

---

<sup>18</sup> Dieses Item lud auf zwei Faktoren recht hoch. Das Argument wird offensichtlich sowohl von Personen gebracht, die genuin an Person B interessiert sind (Faktor Helfen...) als auch von Personen, die hiermit ihren Eigennutz verschleiern möchten (Faktor Eigennutz).

2005). Da zu dieser Frage ex ante keine Hypothesen gebildet wurden, werden die Zusammenhänge hier (explorativ) dargestellt und einige Folgerungen für zukünftige Studien abgeleitet.

Wer hat moralische Bedenken? Personen, die hohe Werte auf  $US_{\text{Nutznießer}}$  aufweisen, aber nicht bestraft haben, äußern in dieser Situation verstärkt moralische Bedenken ( $r = .34$ ;  $p < .01$ ;  $N = 60$ ). Im Gegensatz dazu scheinen die Personen mit hohen Werten auf  $US_{\text{Opfer}}$  kein moralisches Dilemma wahrzunehmen ( $r = -.01$ ;  $p = .47$ ; n. s.;  $N = 60$ ).

Zur Berechnung der Effektstärke  $q$  für Unterschiede zwischen Korrelationskoeffizienten werden diese  $z$ -transformiert und dann die  $z$ -Werte voneinander abgezogen. Für die vorliegende Analyse ergibt sich ein  $q$  von  $.34$ . Die Differenz der Korrelationskoeffizienten entspricht einem mittleren Effekt nach Cohen (1988, kleine Effektstärke:  $q = .10$ ; mittlere Effektstärke:  $q = .30$ ; große Effektstärke:  $q = .50$ ).

Wer würde helfen? Erneut sind es die Versuchspersonen mit den hohen Werten auf  $US_{\text{Nutznießer}}$ , die angeben, sich eher prosozial verhalten zu wollen, denn sie weisen auch die höheren Werte beim Item „Helfen: ja, Bestrafen: nein“ auf ( $r = .24$ ;  $p < .05$ ;  $N = 60$ ). Im Gegensatz dazu deutet sich ein negativer Zusammenhang mit der Dimension  $US_{\text{Opfer}}$  an ( $r = -.14$ ;  $p = .14$  n. s.;  $N = 57$ ). Auch hier zeigt sich wieder, dass sich die Korrelationskoeffizienten signifikant unterscheiden ( $q = .38$ ; mittlere Effektstärke).

Wer zeigt in seinem Kommentar eigennützige Tendenzen? Absolut im Einklang mit früheren Studien zur US, sind es die Versuchspersonen mit hohen Werten auf  $US_{\text{Opfer}}$ , die eine Tendenz zum Eigennutz erkennen lassen ( $r = .25$ ;  $p < .05$ ;  $N = 59$ ). Für die VPn mit hohen Werten auf  $US_{\text{Nutznießer}}$  deutet sich ein gegenteiliger Zusammenhang an, der allerdings nicht signifikant ist ( $r = -.11$ ;  $p = .21$ ;  $N = 60$ ). Dennoch ist die Differenz zwischen den Korrelationskoeffizienten sehr deutlich ( $q = .37$ ; mittlerer Effekt).

Aufgrund ihrer Äußerungen, aber auch aufgrund ihrer Persönlichkeitsstruktur lässt sich also vermuten, dass ein Teil der VPn bereit ist zu helfen, nicht aber zu bestrafen. Ob dem tatsächlich so ist oder ob es sich lediglich um „cheap talk“ handelt, um moralisch zu erscheinen, sollte in zukünftigen Studien empirisch überprüft werden.

Neben der Entscheidung über die Bestrafung wurden die Versuchspersonen auch gefragt, wie sie das Geld selbst aufgeteilt hätten, wenn sie in der Position der Person A gewesen wären. Diese Frage ist eine hypothetische Version der Entscheidung, welche die VPn bei Fetchenhauer und Huang (Fetchenhauer & Huang, 2004) im Diktator-Spiel fällen mussten. Auch hier zeigt sich ein konsistentes Bild: Personen, die hohe Werte auf  $US_{\text{Nutznießer}}$  aufweisen, hätten der anderen Person mehr Geld abgegeben als Personen mit niedrigen Werten auf dieser Dimension ( $r = .21$ ;  $p < .05$ ;  $N = 91$ ; kleine bis mittlere Effektstärke).

### *Diskussion der Studie 2*

Eines der Ziele der Studie 2 war die Replikation der Ergebnisse der Studie 1. Dies ist – zumindest auf bivariater Ebene – gelungen. Es zeigen sich von der Struktur her die gleichen Zusammenhänge wie in Studie 1. Personen mit hohen Werten auf  $US_{\text{Nutznießer}}$  berichten stärkere negative Emotionen und bestrafen eher als Personen mit niedrigen Werten auf  $US_{\text{Nutznießer}}$ . Die Zusammenhänge sind allerdings durchweg niedriger als in Studie 1 und auch der Mediationseffekt deutet sich nur an, ist aber nicht signifikant.

Sehr interessante neue Einblicke lieferte die Analyse der Kommentare derer, die nicht bestraft hatten. Verschiedentlich wurden diejenigen, die kein Third-Party-Punishment betreiben, als Trittbrettfahrer zweiter Ordnung (Bowles & Gintis, 2004; vgl. z.B. Panchanathan & Boyd, 2004) bezeichnet, was wiederum eine Eigennutzmotivation unterstellt. Dass es für ein solches Verhalten allerdings ganz unterschiedliche Motivationen geben kann, zeigen die Analysen der Nicht-Bestrafer. Eine dichotome Trennung in Bestrafen und Nicht-Bestrafen ist nicht zielführend. Hinter dem gleichen beobachtbaren

Verhalten – Nicht-Bestrafen – verbergen sich offensichtlich verschiedene Motive verschiedener Persönlichkeiten. Eine Gruppe war anscheinend von Eigennutzmotiven getrieben. Diese Personen wollten nicht auf die 5 Euro verzichten und bestrafen daher nicht. Eine andere Gruppe bestrafte nicht, weil sie in der Bestrafung keinen Sinn sah. Diese Personen folgen einer Logik, die Mackie (1985, S. 207f) beschreibt. Das Prinzip des „positive retributivism“, wonach ein Schuldiger bestraft werden muss, wird von einigen Personen abgelehnt, wenn sich die Strafe nicht durch eine zu erwartende zukünftige Verhaltensänderung (z.B. durch Abschreckung) des Bestraften rechtfertigen lässt<sup>19</sup>.

Die VPn, die so argumentierten, hätten vielleicht geholfen, indem sie Person B Geld gegeben hätten, sie waren aber nicht bereit, nur zu bestrafen. Diese Behauptungen sind ungeprüft und im Sinne der Ökonomie „cheap talk“, aber der Zusammenhang dieser Aussagen mit der Persönlichkeit ihrer Verfasser ist deutlich. Erneut ist es die prosozial ausgerichtete Dimension  $US_{Nutznießer}$ , die mit der Äußerung, helfen zu wollen, korreliert, und es sind auch Personen mit hohen Werten auf  $US_{Nutznießer}$ , aus deren Kommentaren moralische Bedenken sichtbar werden. Im Gegensatz dazu sind es VPn mit hohen Werten auf der Dimension  $US_{Opfer}$ , die in ihren Begründungen für das Nicht-Bestrafen eigennützige Beweggründe für ihr Verhalten erkennen lassen.

Fetchenhauer und Huang (2004) berichten einen Unterschied in den Mittelwerten der drei Dimensionen der Ungerechtigkeitssensibilität, der darauf hindeutet, dass die Versuchspersonen am stärksten negativ auf eigene erlebte Ungerechtigkeit reagieren. Sie reagieren schwächer sowohl auf beobachtete Ungerechtigkeit als auch, wenn sie Nutznießer sind. Ein solcher Mittelwertsunterschied deutet sich für die Stichprobe von Studie 2 an, ist allerdings nicht signifikant. In Studie 1 sowie in der Untersuchung von Schmitt, Maes und Schmal (1997) zeigt sich kein solcher Unterschied. Dieses Phänomen, das

---

<sup>19</sup> Dieses Verhalten deutet eher auf eine utilitaristische im Gegensatz zu einer deontischen Auffassung von Gerechtigkeit hin.

sich in der unterschiedlichen Bereitschaft zu Second- und Third-Party-Punishment widerzuspiegeln scheint, sollte weiter untersucht werden.

Wieso kam es zu geringerer Bestrafung als in vergleichbaren Untersuchungen (z.B. Fehr & Fischbacher, 2004b)? Ein Grund bestand evtl. darin, dass den VPn das Geld, mit dem sie bestrafen konnten, vor dem Versuch bar ausgehändigt wurde. Üblicherweise werden die Versuchspersonen erst am Ende des Versuchs bezahlt, vorher spielen sie meist um Punkte (vgl. Fehr & Fischbacher, 2004b), die natürlich viel abstrakter sind als reales Geld. Daher ist davon auszugehen, dass das Hergeben des Bargeldes, das sich schon im Besitz der VPn befindet, auch einen stärkeren Strafreiz darstellt als das Zuweisen von (abstrakten) Bestrafungspunkten, die nach dem Experiment zu einer geringeren Auszahlung führen (vgl. Wiswede, 2007, der für abstraktere Geldformen einen verminderten Strafreiz bei der Bezahlung annimmt). Ein weiterer Grund kann in den höheren Bestrafungskosten gesehen werden (siehe oben den Abschnitt zu den Kosten der Bestrafung im theoretischen Teil). Die Kosten in dieser Studie lagen bei 1:2, während sie bei Fehr und Fischbacher 1:3 betrugen, was auch zu einer geringeren Bestrafung geführt haben dürfte.

Die geringere Bestrafungsneigung führte zu einem anderen interessanten Aspekt: Es zeigt sich in dieser Studie deutlich, dass auch bei Third-Party-Punishment bei einem Verhältnis von Kosten zu Stärke der Bestrafung von 1:2 die ungleiche Aufteilung für den Aufteilenden lohnenswert ist. Es bestraften nur 30% der Versuchspersonen, so dass der Erwartungswert des unfairen Angebots (10:0) bei ca. 8.50 Euro liegt. Das ist deutlich mehr als der Erwartungswert eines fairen Angebots, der erfahrungsgemäß bei ziemlich genau 5 Euro liegt. Bei einem Verhältnis von 1:3 in der Studie von Fehr und Fischbacher (2004b) ist der Vorteil schon deutlich geringer. Es könnte also sein, dass sich bereits bei einem Verhältnis von 1:4 die

ungleiche Aufteilung nicht mehr lohnt. Zukünftige Studien sollten diesen Zusammenhang genauer überprüfen.

### **5.1.3. Gesamtdiskussion der Studien 1 und 2**

Die gleichzeitige Betrachtung des Einflusses von negativen Emotionen (Ärger, Empörung) und Persönlichkeitseigenschaften (USNutznießer) auf das (Bestrafungs-) Verhalten in experimentellen Spielen bedeutet die Zusammenführung zweier Forschungsstränge. In der Vergangenheit wurden verschiedentlich Untersuchungen zum Einfluss der negativen Emotionen auf das (Bestrafungs-)Verhalten in experimentellen Spielen gemacht und auch der Einfluss von Persönlichkeitsvariablen wurde untersucht. Eine gemeinsame Betrachtung und die Untersuchung des Einflusses, den die Persönlichkeit auf die Entstehung negativer Emotionen hat, und wie diese dann auf das Verhalten wirken, wurde meines Wissens bisher noch nicht durchgeführt.

Beide Studien weisen die Bereitschaft zu altruistischer Bestrafung (Third-Party-Punishment) eindeutig nach. 40% der VPn, die alle von einer unfairen Aufteilung selbst nicht betroffen waren, lehnten in Studie 1 diese Verteilung dennoch ab, und das, obwohl sie dafür auf 5 Euro verzichten mussten. In Studie 2 waren immerhin ca. 30% der Versuchspersonen bereit, Geld dafür zu bezahlen, dass eine unfaire fremde Person bestraft wird, indem ihr Geld abgezogen wird.

Beide Studien konnten zudem eindeutig einen Einfluss der Ungerechtigkeitssensibilität auf die Bestrafung nachweisen. Insofern gliedern sie sich nahtlos in eine Reihe von Studien zu experimentellen Spielen und US ein (Fetchenhauer & Huang, 2004). Außerdem sind sie im Einklang mit verschiedenen Studien, die einen Zusammenhang der US mit pro- und antisozialem Verhalten nachweisen (Gollwitzer et al., 2005).

Ein weiterer wichtiger Befund der Studien 1 und 2 ist der Einfluss negativer Emotionen auf die Bestrafung: Es konnte gezeigt werden,

dass der Bestrafung verstärkt negative Emotionen vorausgingen. Wie bereits weiter oben geschildert, wird den Emotionen ein unterschiedlich hoher Einfluss auf altruistische Bestrafung zugesprochen.

Die Verbindung der negativen Emotionen in als ungerecht wahrgenommenen Situationen zur US lässt sich direkt aus den theoretischen Überlegungen zur US ableiten (s. o. den theoretischen Abschnitt zur US). Hier wurde dieser Zusammenhang erstmalig für experimentelle Spielsituationen empirisch nachgewiesen. Bemerkenswert ist der lange Zeitraum zwischen der Erfassung der US und dem Experiment in Studie 2. Zwischen der Erfassung des Persönlichkeitsmerkmals US und dem Experiment vergingen etwa vier Wochen. Hierauf dürfte der geringere Zusammenhang zwischen der US und dem Bestrafungsverhalten sowie zwischen der US und den negativen Emotionen im Vergleich zur Studie 1 zurückzuführen sein.

Die Tatsache, dass der Mediationseffekt in beiden Studien nicht signifikant nachgewiesen wurde, deutet darauf hin, dass der Einfluss der Ungerechtigkeitssensibilität nicht nur über moralische Emotionen, sondern auch über andere Variablen vermittelt wird. Verschiedentlich wurde nachgewiesen, dass der Einfluss der US auch über moralische Kognitionen auf das Bestrafungsverhalten wirkt (Gollwitzer et al., in Druck; Baumert, 2003).

Der Anteil der Versuchspersonen, die bestraft haben, ist in beiden Studien niedriger als z.B. bei Fehr und Fischbacher (2004b). Dies kann verschiedene Gründe haben:

1. Höhere Kosten der Bestrafung: In beiden Studien waren die Kosten der Bestrafung höher als bei Fehr und Fischbacher (Studie 1: Kosten fix, Verhältnis 5/9; Studie 2: Kosten variabel, Verhältnis 1:2; Fehr und Fischbacher: Verhältnis 1:3).
2. Keine Abstufungen der Bestrafung: In Studie 1 war eine graduelle Bestrafung nicht möglich. Eine Entscheidung zwischen Annahme oder Ablehnung der Entscheidung bei fixen Kosten von 5 Euro musste gefällt werden.

3. Bargeld statt Punkte: Wie oben bereits erwähnt, stellt das Herausgeben von Geld in Münzform einen höheren Strafreiz dar als das Vergeben von Punkten.

Das Erfassen der Versuchspersonenkommentare stellte sich aus verschiedenen Gründen als sehr fruchtbar heraus. Hieraus konnten die Emotionen der Versuchspersonen ermittelt werden, ohne diese dabei zu beeinflussen. Darüber hinaus war es möglich, die Motive der Versuchspersonen, die nicht bestraft hatten, genauer zu untersuchen. Ein methodischer Nebenaspekt der Kommentare ist, dass auf diese Weise VPn identifiziert werden können, die trotz korrekt gelöster Kontrollaufgaben die experimentelle Situation anscheinend nicht verstanden hatten.

Die Untersuchung derer, die nicht bestraft, scheint mir ein sehr wichtiger Punkt zu sein. Meist wird in experimentellen Spielen zu einseitig auf Bestrafung zur Stärkung von Normen der Kooperation und des Altruismus eingegangen.<sup>20</sup> Transfers und Hilfeleistungen kommen in der realen Welt vor und daher sollten sie auch entsprechend in experimentellen Spielen modelliert werden. Es sollte aber kein reines „entweder / oder“ sein, sondern es sollten alle Möglichkeiten in einer Untersuchung angeboten werden. In einem solchen Design würde auch vermieden, dass VPn sich zum Bestrafen gedrängt fühlen, weil dies die einzige Handlung ist, mit der sie ihre Missbilligung der Ungerechtigkeit ausdrücken können. Hogan & Emler (1981) verweisen allerdings darauf, dass die meisten Personen ohnehin eher bereit sind, Täter zu bestrafen, als Opfern eine Kompensation für deren Schaden zuzusprechen. Dennoch ist es gut möglich, dass die Wahl für eine der beiden Verhaltensweisen von der Persönlichkeit abhängt. Die Voraussagekraft aufgrund der Persönlichkeitsvariablen wäre in einem solchen Design vermutlich noch höher als in den hier berichteten Studien 1 und 2, da die VPn eher die Möglichkeit hätten,

---

<sup>20</sup> Das drückt sich auch in der Bezeichnung „altruistische Bestrafung“ aus. In der Einleitung wurde bereits auf „norm enforcement“ und „strong reciprocity“ verwiesen, die Strafen und Belohnungen einschließen.



sich entsprechend ihrer Persönlichkeit zu verhalten. Erste Hinweise hierauf liefert folgende Auswertung: Wenn man in Studie 2 diejenigen VPn aus der Analyse ausschließt, die angaben, helfen zu wollen, aber nicht bestrafen, steigt der Zusammenhang der US mit der Bestrafung an  $r = .21^*$  auf  $r = .36^{**}$  signifikant an ( $p < .01$ ;  $N = 56$ ).

Ein Design, das sowohl Bestrafung als auch Belohnung und Transfer in Abhängigkeit von der Ungerechtigkeitssensibilität der Versuchspersonen untersucht, wäre eine interessante Erweiterung besonders der zweiten Studie. Ein ähnlicher Versuch, allerdings ohne die Anbindung an die Emotion und Persönlichkeit der Versuchspersonen sowie ohne die Möglichkeit eines reinen (Zwangs-)Transfers vom Diktator an den Empfänger (ohne weitere Bestrafung) wurde von Ottone (2005) bereits durchgeführt. In einem erweiterten Design (u. a. mit der Möglichkeit für die dritte Person, dem Empfänger eigenes Geld zu geben oder einen Zwangstransfer vom Aufteiler an den Empfänger anzuordnen), ist zu erwarten, dass die US zwischen verschiedenen Verhaltensweisen diskriminieren kann. Wie in Studie 2 gezeigt wurde, liegen dem Nicht-Bestrafen sowohl altruistische als auch eigennützige Motive zu Grunde. Deshalb kann vermutet werden, dass eigennützige Versuchspersonen (hohe Werte auf  $US_{\text{Opfer}}$ ) in dem erweiterten Design nicht bestrafen, nicht helfen und auch nichts transferieren. Von VPn mit hohen Werten auf  $US_{\text{Nutznießer}}$  kann erwartet werden, dass sie entweder bestrafen oder aber helfen, indem sie Geld transferieren. Täten sie dies nicht, wären ihre Bekundungen tatsächlich nichts weiter als „cheap talk“ gewesen.

## **5.2. Die Wirksamkeit von altruistischer Bestrafung: Berücksichtigen Individuen die Bestrafungsneigung ihrer Mitmenschen?<sup>21</sup>**

The impulse to seek revenge is likewise counterproductive unless others have some way of anticipating that one has it.

(Frank, 1988, S. 54)

Berücksichtigen Menschen, dass andere altruistisch bestrafen? Die Vermutung, Menschen könnten antizipieren, dass unfair behandelte sich wehren, liegt nahe und wurde für Second-Party-Punishment verschiedentlich geäußert und untersucht (vgl. z.B. Fehr & Fischbacher, 2002; Hoffman et al., 1996). Wesentlich interessanter ist die Frage, wie es mit Third-Party-Punishment durch eine unbeteiligte dritte Person aussieht. Würden Menschen diese Art der Bestrafung nicht antizipieren, hätte sie keinen Einfluss auf ihr Verhalten und somit auch keinen positiven Ertrag (zukünftige Normeinhaltung) für die Gesellschaft. Wird aber die Bestrafung erwartet, bedeutet das, dass die Bestrafung durch Dritte Kooperation und Normeinhaltung fördert. Schon allein die Androhung der Bestrafung hätte dann einen abschreckenden Charakter und würde deutlich zur Effizienz des Bestrafungssystems beitragen. Je seltener die Bestrafung tatsächlich durchgeführt werden muss, desto effizienter ist das System, da bei Bestrafung immer Kosten sowohl für den Bestrafenden als auch für den Bestraften entstehen. Die Annahme der Abschreckung ist ein Hauptbestandteil der Logik des rechtlichen Strafsystems westlicher Gesellschaften.

Der Frage der Antizipation wurde bisher nur sehr wenig Aufmerksamkeit geschenkt. Zwar bestätigten Vergleiche zwischen Ultimatum- und Diktatorspiel schon recht früh die Vermutung, dass ein großer Teil des scheinbar großzügigen Verhaltens im Ultimatum-Spiel in Wirklichkeit auf strategische Erwägungen zurückgeführt werden

---

<sup>21</sup> Auf den gleichen Untersuchungen beruht auch ein noch nicht publizierter Artikel von Fetschenhauer, Jacobs und Gresser.

muss (vgl. z.B. Forsythe et al., 1994), zur Frage der Antizipation von altruistischem Bestrafen durch Dritte wurde bisher allerdings recht wenig geforscht. Eine Ausnahme bilden Bernhard und Fehr et al. (2006), die explizit nach Bestrafungserwartungen fragen und einen positiven Zusammenhang der Bestrafungserwartung mit dem transferierten Betrag im Diktator-Spiel finden. Fehr und Fischbacher (2004b) behandeln diese Frage lediglich indirekt. Sie berichten, dass in ihrer ersten Studie die Angebote an den Dummy im Third-Person-Punishment-Spiel denen aus einfachen Diktatorspielen glichen. Bedeutet das, dass sich deren Aufteilende nicht von der Bestrafungsmöglichkeit der dritten Partei beeinflussen ließen, wenn sie ungleiche Angebote abgaben? Wären ihre Aufteilungen die gleichen gewesen, wenn sie an einem einfachen Diktator-Spiel teilgenommen hätten?

Die Hauptforschungsfragen, die mit den Studien 3 und 4 beantwortet werden sollen sind:

1. Sehen die Aufteilenden altruistische Bestrafung voraus und ändert sich deshalb ihr Aufteilungsverhalten gegenüber einer machtlosen anderen Person, wenn sie wissen, dass sie durch eine dritte Partei für selbstsüchtiges Verhalten bestraft werden können?
2. Sind die Aufteilenden in der Lage, zwischen der unterschiedlichen Bereitschaft zu (altruistischer) Bestrafung einer selbst vom unfairen Akt betroffenen Person (Second-Party-Punishment) und einer nicht betroffenen dritten Person (Third-Party-Punishment) zu unterscheiden?
3. Wenn die Aufteilenden die unterschiedliche Bereitschaft zu Second- und Third-Party-Punishment richtig antizipieren können, reagieren sie dann strategisch darauf?
4. Wenn die Aufteilenden strategisch reagieren, welche Strategie wirft dann den höchsten Gewinn für einen Aufteilenden ab?

Bei der Auswertung der Strategien wird zwischen verschiedenen Heuristiken (vereinfachte Lösungsansätze, „Daumenregeln“) unterschieden, die Menschen anwenden könnten, wenn sie an einem spieltheoretischen Experiment teilnehmen. Grundsätzlich kann man eine solche Situation von einem ethischen (Was ist eine faire Aufteilung in dieser Situation?) oder von einem strategischen Standpunkt (Was ist die beste Strategie um meinen Gewinn zu maximieren?) aus betrachten. Eine Möglichkeit, in Erfahrung zu bringen, nach welcher Heuristik sich die Versuchspersonen wohl verhalten, ist der Einsatz von Persönlichkeitsmaßen, wie in Studie 1 und 2. Eine andere – weniger aufwändige – Methode wurde hier verfolgt: Wie oben verschiedentlich erwähnt, dient das Diktator-Spiel dazu, Altruismus zu messen. Dementsprechend wurde in den Studien 3 und 4 der Unterschied in fairem Verhalten über ein Diktator-Spiel gemessen. Im Diktator-Spiel gibt es aufgrund der fehlenden Sanktionsmöglichkeiten durch den Empfänger für den Aufteiler keinen strategischen Anreiz zu einer Gleichverteilung. Deshalb wurden jene Personen, die dennoch eine Gleichverteilung vornahmen, als ethisch bzw. fairnessorientiert bezeichnet. Diejenigen, die mehr als die Hälfte für sich behielten, wurden entsprechend als zumindest teilweise von eigennützigen Interessen angetrieben klassifiziert. Im anschließenden Drei-Personen-Ultimatum-Spiel wird untersucht, wie gut diese beiden Probanden-Typen einschätzen können, wie wahrscheinlich eine Bestrafung durch den Spielpartner sein würde.

### **5.2.1. Studie 3**

Zur Messung altruistischer Bestrafung wurde ein Drei-Personen-Ultimatum-Spiel – Aufteiler, Entscheider, Dummy– eingesetzt (ähnlich dem von Güth & van Damme, 1998). Das Spiel fand unter anonymen Bedingungen statt, die drei Personen lernten sich also weder während noch nach dem Spiel kennen. Die Aufteiler erhielten vom Versuchsleiter einen bestimmten Geldbetrag, den sie zwischen sich, dem Entscheider und dem Dummy aufteilen sollten. Die Aufteiler mussten

also angeben, wie viel Geld sie selbst behalten und wie viel sie an den Entscheider und an den Dummy geben wollten. Nachdem die Aufteiler ihre Entscheidung getroffen hatten, wurden die Entscheider darüber informiert und hatten dann zwei Möglichkeiten: entweder sie akzeptierten die Aufteilung des Aufteilers, dann erhielt jeder den vom Aufteiler vorgesehenen Betrag oder sie lehnten ab, dann erhielten alle drei nichts. Zwei Fragen stehen im Mittelpunkt dieser Studie: Antizipieren die Aufteiler, dass ihre Entscheidung unter Umständen abgelehnt wird, wenn sie zu gierig sind? Können Sie voraussehen, dass die Bereitschaft zu Third-Party-Punishment evtl. geringer ist als die Bereitschaft zu Second-Party-Punishment?

Diese Fragen lassen sich beantworten, indem man die Angebote an den Dummy im Drei-Personen-Ultimatum-Spiel mit den Angeboten an den machtlosen Empfänger im Diktator-Spiel vergleicht. Sind die Angebote an den Dummy im Drei-Personen-Ultimatum-Spiel höher als an den machtlosen Empfänger im Diktator-Spiel, deutet das darauf hin, dass die Aufteilenden tatsächlich in Betracht ziehen, für die Ausnutzung der Machtposition bestraft zu werden.

Außerdem wird die Frage untersucht, welche Strategie den größten Gewinn verspricht, d.h., welche Aufteilungsstrategie unter Berücksichtigung der Bestrafungsneigung der Entscheider für den Aufteilenden am lukrativsten ist.

### *Methode*

Die Teilnehmer der Studie waren 190 Studenten der Ruhruniversität Bochum. 60% von ihnen studierten Psychologie und 40% Recht. Der Anteil an Frauen lag bei 71% und das durchschnittliche Alter der Probanden betrug 24 Jahre. 9 Personen mussten von der Analyse ausgeschlossen werden, da sie entweder nicht alle Angaben gemacht hatten oder ihre Antworten zeigten, dass sie die verwendeten spieltheoretischen Paradigmen nicht verstanden hatten. Aus diesem Grund beziehen sich alle Analysen auf N=181 Versuchspersonen.

Zunächst wurden den Probanden die Regeln der beiden Spiele über schriftliche Instruktionen erläutert (Diktator-Spiel, Drei-Personen-Ultimatum-Spiel). Den Versuchspersonen wurde erklärt, dass sie über einen Zufallsprozess den Rollen in den Spielen zugewiesen würden (sie wussten *ex ante* also z.B. nicht, ob sie Aufteiler oder Empfänger im Diktator-Spiel sein würden).

Im Diktator-Spiel wurden 100 Geldeinheiten verteilt, im Drei-Personen-Ultimatum-Spiel 150 Geldeinheiten. Eine Geldeinheit entsprach ungefähr 0.60 Euro, so dass es im Diktator-Spiel also um ca. 60 Euro und im Drei-Personen-Ultimatum-Spiel insgesamt um ca. 150 Euro ging.

Den Probanden wurde mitgeteilt, dass ihre Entscheidungen nicht hypothetisch seien, denn jede 20. Entscheidung werde per Lotterie ausgewählt und führe zu echten Auszahlungen entsprechend den Entscheidungen. Fehr und Schmidt (1999) konnten zeigen, dass das Verhalten nicht davon abhing, ob jede einzelne Entscheidung zu Auszahlungen führt oder ob nur ein Teil der Entscheidungen durch einen Zufallsmechanismus auszahlungsrelevant wird. Bevor die Probanden ihre Entscheidungen fällten, wurden sie darauf hingewiesen, dass im Drei-Personen-Ultimatum-Spiel niemand etwas erhält, wenn das Angebot des Aufteilenden unter dem Betrag liegt, der als minimaler Akzeptanzbetrag vom zugelosten Entscheider angegeben wurde. Jeder Proband musste sowohl für das Diktator-Spiel als auch für das Drei-Personen-Ultimatum-Spiel eine Aufteilung des Geldbetrages als Aufteilender vornehmen. Darüber hinaus mussten alle Probanden auch angeben, welchen Betrag sie als Entscheider mindestens erhalten wollten und was der Dummy mindestens erhalten müsste, damit sie im Drei-Personen-Ultimatum-Spiel die Aufteilung akzeptierten. Die Angabe der Beträge erfolgte ohne jede Vorgabe, d.h. die Probanden trugen die Summe, die sie anbieten bzw. mindestens für sich und den Dummy erhalten wollten, in ein dafür vorgesehenes leeres Feld ein. Um die Anonymität der Versuchspersonen zu gewährleisten, trug

jeder Proband auf seinen Entscheidungsbogen ein von ihm selbst zu generierendes Codewort ein. Über diese Codewörter erfolgte die Zuordnung der verschiedenen Entscheidungsbögen und der späteren Bezahlung.

Die Beschreibung der Spiele war wie immer neutral gehalten, um keine Reaktionen der Versuchspersonen in eine bestimmte Richtung hervorzurufen.

Die Reihenfolge, in der die Probanden ihre Angaben zum Diktator-Spiel bzw. zum Drei-Personen-Ultimatum-Spiel machen mussten, war über die gesamte Stichprobe ausgeglichen, d.h. die Hälfte erhielt zunächst das eine Spiel und dann das andere, bei der anderen Hälfte war es umgekehrt. Die Reihenfolge hatte keinen signifikanten Einfluss auf die Entscheidungen der Probanden und wird daher nicht weiter diskutiert.

Eine Woche nach der Untersuchung wurden alle Teilnehmer über den Versuch aufgeklärt (debriefing) und jeder 20. wurde entsprechend der eigenen und der Entscheidung eines zufällig ausgewählten Interaktionspartners entweder im Diktator-Spiel oder im Drei-Personen-Ultimatum-Spiel bezahlt. Die Versuchspersonen erhielten ihre Bezahlung in einem verschlossenen Umschlag, auf dem ihr Codewort stand.

### *Ergebnisse*

Wie haben die Versuchspersonen das Geld im Drei-Personen-Ultimatum-Spiel, in dem es um 150 Geldeinheiten ging, verteilt? Die Aufteilenden teilten sich im Durchschnitt 55.3 Geldeinheiten zu ( $SD = 12.5$ ) und dem Entscheider und dem Dummy zusammen 94.7 Geldeinheiten. Ein Großteil, nämlich 77.5%, der Aufteilenden nahm sich exakt ein Drittel (50 Geldeinheiten) des aufzuteilenden Betrags, wohingegen 22.5% der Versuchspersonen mehr als ein Drittel für sich beanspruchten.

Die Aufteilenden boten dem Entscheider durchschnittlich 49.9 Geldeinheiten ( $SD = 6.9$ ) an, also ziemlich genau ein Drittel. Eine Minderheit von 10.6% der Aufteilenden offerierte dem Entscheider weniger als 50 Geldeinheiten, 9.7% boten mehr als 50 Geldeinheiten an. Das durchschnittliche Angebot für den Dummy lag bei 44,8 Geldeinheiten ( $SD = 11.8$ ). Die Mehrheit der Aufteilenden (78%) bot dem Dummy ungefähr ein Drittel an. Die anderen gaben entweder weniger (19.8%) oder sogar gar nichts (2.2%).

Im Durchschnitt waren die Angebote an den Entscheider 5.19 Geldeinheiten höher als an den Dummy ( $t = 4.750$ ;  $p < .01$ ). Die Aufteilenden gehen also anscheinend davon aus, dass die Entscheider eher ein geringeres Angebot für den Dummy akzeptieren als für sich selbst.

Welche Aufteilungsmuster ließen sich beobachten? Über drei Viertel (77.5%) der Aufteiler nahmen eine Gleichverteilung vor. Sie gaben 50 Geldeinheiten an den Entscheider und 50 an den Dummy und teilten somit den gesamten Betrag gleichmäßig unter allen drei Beteiligten auf. Zwei andere Muster sahen eine Ungleichverteilung vor: Eine Gruppe von Versuchsteilnehmern (9.6%) behielt mehr als ein Drittel für sich und bot dem Entscheider mehr als ein Drittel, dem Dummy dagegen deutlich weniger. Diese Probanden scheinen also zu versuchen, den Entscheider zu bestechen und hoffen so damit durchzukommen, dass sie dem Dummy fast nichts anbieten. Eine dritte Strategie, die von einigen der Probanden (8%) verfolgt wurde, war, selbst mehr als ein Drittel zu behalten und dem Entscheider und dem Dummy den gleichen geringen Anteil anzubieten.

Wie verhielten sich die Probanden im Diktatorspiel bei der Aufteilung der 100 Geldeinheiten? Hier war der durchschnittliche Betrag, der dem Empfänger angeboten wurde, 38.5 Geldeinheiten ( $SD = 16.3$ ). Während 58.8% der Probanden 50/50 aufteilten, gaben die übrigen 41.2% der Probanden weniger als 50% an den Empfänger und 6% gaben nichts.



Wie vermutet, lagen die Angebote an den Empfänger im Diktator-Spiel signifikant unter denen an den Dummy im Drei-Personen-Ultimatum-Spiel. Die durchschnittliche Differenz zwischen den Angeboten in den beiden Spielen betrug 6.2 Geldeinheiten ( $t = 6.081$ ;  $p < .01$ ).

Der Unterschied zwischen Angeboten an den Empfänger im Diktator-Spiel und den Dummy im Drei-Personen-Ultimatum-Spiel weist einen starken Zusammenhang zu den Entscheidungen der Aufteilenden im Diktator-Spiel auf ( $r = -.70$ ;  $p < .01$ ). Je weniger ein Aufteilender dem Empfänger im Diktatorspiel abgab, umso höher war die Differenz zwischen dem Angebot an den Dummy im Drei-Personen-Ultimatum-Spiel und dem Angebot an den Empfänger im Diktator-Spiel. Zur Verdeutlichung: Selbst die 11 Personen, die im Diktatorspiel alles für sich behielten, boten im Drei-Personen-Ultimatum-Spiel dem Dummy im Durchschnitt 42.3 Geldeinheiten ( $SD = 16.03$ ) an. Da in beiden Spielen von diesen Positionen kein Veto oder Eingreifen möglich ist, lässt sich dieser Unterschied nur dadurch erklären, dass die Aufteilenden im Drei-Personen-Ultimatum-Spiel eine Ablehnung durch den Entscheider fürchteten, wenn sie dem Dummy zu wenig anbieten.

Bei der nächsten Analyse wird die Stichprobe in zwei Gruppen aufgeteilt: 1. Diejenigen, die im Diktator-Spiel dem Empfänger die Hälfte gaben, und 2. jene, die weniger als die Hälfte gaben. Wie oben bereits ausgeführt, wurde hier davon ausgegangen, dass die erste Gruppe auch im Drei-Personen-Ultimatum-Spiel an Fairness und einer gleichen Verteilung interessiert sein dürfte, während von der zweiten Gruppe zu erwarten ist, dass sie sich dieser Aufteilung strategisch nähert.

Genau dieses Ergebnis zeigte sich: Diejenigen, die im Diktator-Spiel gleich aufteilten, gaben auch im Drei-Personen-Ultimatum-Spiel dem Entscheider und dem Dummy den gleichen Betrag. Der durchschnittliche Unterschied zwischen den Angeboten für den Entscheider ( $M = 50.09$ ;  $SD = 2.41$ ) und den Dummy ( $M = 49.81$ ;  $SD = 1.92$ ) betrug .28 Geldeinheiten. Der Unterschied ist nicht signifikant ( $t = 1.000$ , n. s.).

Im Gegensatz dazu boten die Versuchspersonen, die sich im Diktator-Spiel nicht fair verhalten hatten, im Drei-Personen-Ultimatum-Spiel dem Entscheider ( $M = 49.73$ ) deutlich mehr Geld an als dem Dummy ( $M = 37.53$ ). Die Differenz von 12.20 ist hoch signifikant ( $t = 5.064$ ,  $p < .01$ ).

Das analoge Ergebnis zeigt sich, wenn man die Angebote an den Dummy mit denen an den Empfänger im Diktator-Spiel vergleicht: Diejenigen, die eine Gleichverteilung im Diktator-Spiel gewählt hatten, boten auch dem Dummy nahezu ein Drittel an ( $M = 49.81$ ). Der Unterschied von .19 ist nicht signifikant ( $t = 1.000$ ; n. s.). Im Gegensatz dazu boten diejenigen, die im Diktator-Spiel ungleich aufgeteilt hatten, dem Dummy deutlich mehr an ( $M = 37.53$ ) als dem Empfänger im Diktator-Spiel ( $M = 22.21$ ). Dieser Unterschied zwischen den Mittelwerten in den beiden Spielen ist hoch signifikant ( $t = 7.48$ ;  $p < .01$ ).

Die Frage nach der gewinnmaximierenden Strategie ist für das Diktator-Spiel schnell beantwortet. Weil es sich hier um ein anonymes, einmaliges Spiel handelt, besteht sie darin, alles für sich zu behalten und nichts abzugeben. Da dies eine mathematische Wahrheit ist, bedarf es keiner weiteren Erklärung oder empirischer Belege.

Die interessantere Frage ist: Mit welcher Strategie verdienen die Versuchspersonen im Drei-Personen-Ultimatum-Spiel am meisten? Diese Frage lässt sich nicht aus dem Spiel ableiten, sondern hängt von den Beträgen ab, die der Entscheider für sich und für den Dummy mindestens verlangt. Man muss sich also die minimalen Akzeptanzschwellen der Entscheider ansehen: Wie viel muss der Aufteiler dem Entscheider und dem machtlosen Dummy mindestens anbieten, damit der Entscheider nicht ablehnt?

Durchschnittlich verlangten die Entscheider für sich 43.8 Geldeinheiten ( $SD = 12.7$ ). Während fast zwei Drittel der Entscheider (65.4%) angaben, Angebote unterhalb von 50 Geldeinheiten abzulehnen, verlangten 2.1% der Probanden sogar mehr als 50 Geldeinheiten.

Eine nähere Betrachtung dieser Versuchspersonen zeigte, dass sie davon ausgingen, dass der Aufteilende das Geld nur mit ihnen teilt und der Dummy leer ausgehen sollte. Eine kleine Minderheit von 1.1% der VPn, wollte alle Verteilungen annehmen, d.h., sie gaben eine Akzeptanzschwelle von Null Geldeinheiten an.

Die angegebenen Akzeptanzschwellen für den Dummy waren sehr heterogen: Die Mehrheit der Probanden (55.2%) akzeptierten nur Aufteilungen, bei denen der Dummy ein Drittel des Betrages (also 50 Geldeinheiten) erhält. Eine Minderheit von 12.3% hätte jedes Angebot für den Dummy akzeptiert, d.h., die Akzeptanzschwelle lag bei Null Euro. Der durchschnittliche minimale Akzeptanzbetrag für den Dummy betrug 36.2 Geldeinheiten (SD = 18.8). Für sich selbst ( $M = 43.8$ ) verlangten die Entscheider also signifikant mehr als für den Dummy ( $M = 36.2$ ). Die Differenz zwischen den minimalen Akzeptanzschwellen ist hoch signifikant ( $t = 5.886$ ;  $p < .01$ ).

Eine simultane Betrachtung der Akzeptanzschwellen gibt Aufschluss über verschiedene Strategien: Die Mehrheit (51.4%) der VPn war nur bereit, eine Strategie zu akzeptieren, bei der der Geldbetrag gleichmäßig auf alle drei Personen aufgeteilt wurde. Ein weiterer Teil der VPn (15.5%) akzeptierte Angebote, in denen der Dummy und sie selbst zwar jeweils das gleiche, allerdings weniger als ein Drittel des Gesamtbetrages erhielten. Somit gaben zwei Drittel (15.5% plus 51.4%) der VPn für sich und den Dummy die gleiche minimale Akzeptanzschwelle an. 17.4% der Probanden gab für sich eine höhere Akzeptanzschwelle als für den Dummy an, für den die sie aber immer noch mehr als Null Geldeinheiten forderten. Während eine kleinere Gruppe von Probanden (12.3%) für den Dummy eine Akzeptanzschwelle von Null angab, für sich aber eine höhere, gaben 4% der Versuchspersonen für den Dummy eine höhere Akzeptanzschwelle an als für sich und verlangten für den Dummy exakt ein Drittel des Gesamtbetrages.

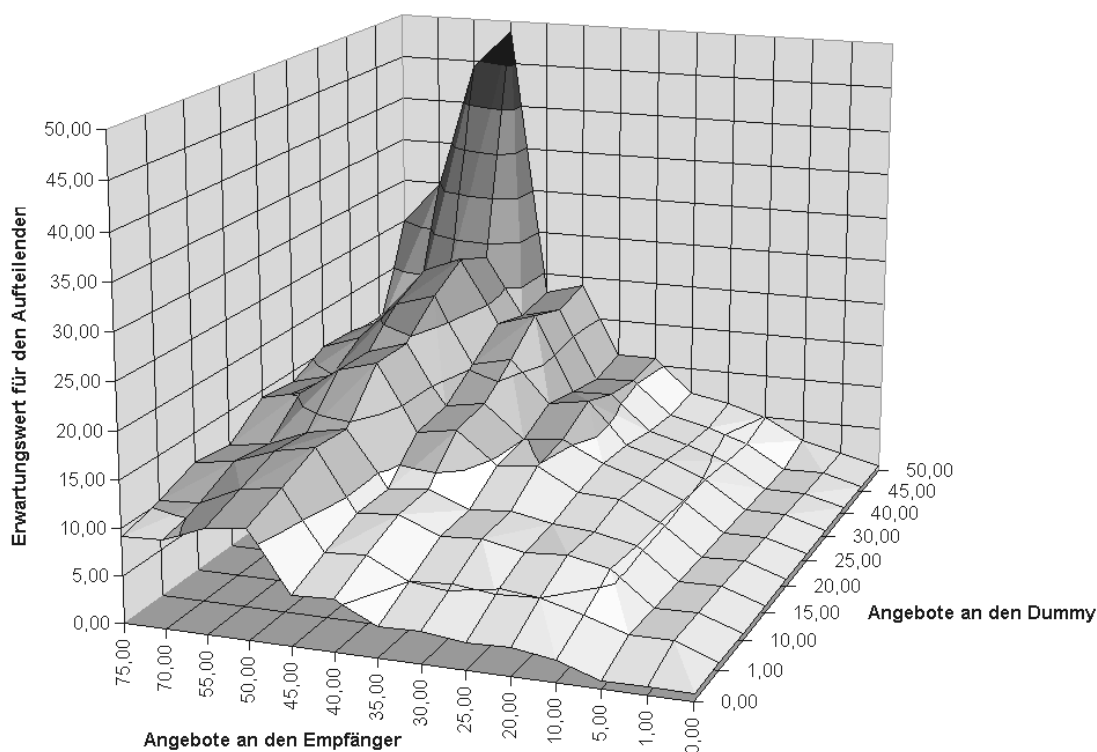
Der Tabelle lassen sich die Wahrscheinlichkeiten entnehmen, mit der ein bestimmtes Angebot akzeptiert wurde (Zeilen und Spalten, in denen keine Angebote vorkamen, wurden nicht aufgenommen). Die Kombination, die mit Abstand die größte Wahrscheinlichkeit akzeptiert zu werden aufweist, ist die Gleichverteilung (50 – 50 – 50) des Geldes. Nahezu alle Personen (98%) gaben an, diese Kombination zu akzeptieren. Eine Aufteilung, die für den Aufteilenden 120 Geldeinheiten, den Entscheider 30 Geldeinheiten und den Dummy 0 Geldeinheiten vorsehen würde, wäre z.B. nur von 3% der Entscheider akzeptiert worden und selbst der „Bestechungsversuch“ (75 für den Aufteilenden und den Entscheider und 0 für den Dummy) hätte nur in 12% der Fälle gefruchtet.

		Angebote an den Dummy									
		0	1	10	15	20	25	30	40	45	50
Angebote an den Entscheider	0	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	1	0,01	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
	5	0,01	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,03
	10	0,02	0,03	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,06
	20	0,02	0,04	0,06	0,06	0,07	0,07	0,07	0,07	0,07	0,08
	25	0,02	0,04	0,06	0,06	0,07	0,08	0,08	0,08	0,08	0,09
	30	0,03	0,04	0,07	0,07	0,10	0,11	0,14	0,14	0,14	0,15
	35	0,03	0,04	0,07	0,07	0,10	0,11	0,14	0,14	0,14	0,17
	40	0,05	0,07	0,09	0,09	0,14	0,16	0,20	0,28	0,28	0,31
	45	0,05	0,07	0,09	0,09	0,14	0,16	0,20	0,28	0,28	0,32
	50	0,12	0,13	0,17	0,17	0,24	0,26	0,33	0,42	0,43	<b>0,98</b>
	55	0,12	0,13	0,17	0,17	0,24	0,26	0,33	0,43	0,43	0,98
	70	0,12	0,13	0,18	0,18	0,25	0,27	0,34	0,44	0,44	0,99
	75	0,12	0,14	0,18	0,19	0,26	0,28	0,35	0,44	0,45	1,00

**Tabelle 2: Wahrscheinlichkeit mit der ein Angebot akzeptiert wurde**

Aus der Tabelle mit den Wahrscheinlichkeiten in Kombination mit den Beträgen, die ein Aufteilender für sich vorgesehen hat, kann man den Erwartungswert für alle möglichen Kombinationen von Angeboten an den Entscheider und den Dummy errechnen. Nachdem die Wahrscheinlichkeiten schon bekannt sind und der Erwartungswert eine multiplikative Verknüpfung des Geldbetrages für den Aufteilenden mit der Wahrscheinlichkeit ist, verwundert das Ergebnis nicht

sonderlich: Die gewinnmaximierende Strategie für den Aufteilenden – unter Berücksichtigung der Ablehnungs- bzw. Bestrafungsneigung seiner Mitspieler – ist die faire Strategie (50 - 50 - 50). Diese Strategie resultiert in einem Erwartungswert von ca. 49 Geldeinheiten. Abbildung 9 verdeutlicht den Verlauf der zu erwartenden Gewinne für die möglichen Angebote. Dieses Ergebnis reiht sich gut in eine Vielzahl von Studien ein, die zeigen, dass die Gleichverteilung des Geldes meist auch die gewinnmaximierende Strategie ist (s. o.). Selbst die zweitbeste Strategie (50 für den Entscheider und 40 für den Dummy) führte schon nur noch zu einem Erwartungswert von ca. 25 Geldeinheiten – knapp der Hälfte also. Der weiter oben bereits angesprochene „Bestechungsversuch“ (75 - 75 - 0) verfehlte sein Ziel völlig und führte lediglich zu einem Erwartungswert von ca. 9 Geldeinheiten.



**Abbildung 9: Erwartungswert der Aufteilung für den Aufteilenden in Abhängigkeit vom Angebot an den Entscheider und den Dummy im Drei-Personen-Ultimatum-Spiel**

Je mehr Geld die Aufteilenden anboten, desto höher war ihr zu erwartendes Ergebnis ( $r = .93$ ;  $p < .01$ ). Interessanterweise lehnten die Entscheider sowohl Angebote ab, die für sie weniger als 50 Geldeinheiten vorsahen als auch solche, die mehr als 50 Geldeinheiten einbrachten.

Weiter oben wurde differenziert zwischen denen, die im Experiment fair waren und sich streng an eine Gleichverteilungsnorm hielten, und denen, die eher strategisch vorgingen. Über den Erfolg dieser beiden Gruppen lässt sich ein klares Urteil fällen: Die „Strategen“, die im Diktatorspiel weniger als 50 Geldeinheiten abgaben, verdienten im Drei-Personen-Ultimatum-Spiel im Durchschnitt 32.9 Geldeinheiten ( $SD = 15.69$ ). Diejenigen, die bei beiden Spielen einfach eine Gleichverteilung vornahmen, verdienten 48.7 Geldeinheiten ( $SD = 3.61$ ), also nahezu 50% mehr ( $t = 7.980$ ;  $p < .01$ ).

### *Diskussion*

Die Frage, die diese Studie zu klären versuchte, ist: Antizipieren die Aufteiler die altruistische Bestrafung unfairer Angebote durch die Entscheider? Die Antwort lautet: Ja! Die Versuchspersonen sind in der Lage, zwischen den unterschiedlichen Situationen zu unterscheiden. Sie gaben mehr Geld an den Entscheider im Drei-Personen-Ultimatum-Spiel als an den Empfänger im Diktator-Spiel. Sie antizipieren also, dass diese Versuchspersonen sich wehren, wenn sie die Möglichkeit dazu haben. Auf der anderen Seite gaben die Aufteilenden aber auch mehr Geld an die Entscheider im Drei-Personen-Ultimatum-Spiel als an die Dummies. Sie sehen also voraus, dass die Bereitschaft zu Third-Party-Punishment geringer ist als zu Second-Party-Punishment (vgl. auch Fehr & Fischbacher, 2004b). Dennoch erkannten sie aber, dass es eine gewisse Bereitschaft zu Third-Party-Punishment gibt, denn sie gaben den Dummies mehr als den Empfängern im Diktator-Spiel.

Ein Teil der Versuchspersonen unterschied also strategisch zwischen den Situationen, aber die Analyse der Erwartungswerte zeigt, dass

diese Diskrimination nicht zu höheren, sondern zu niedrigeren Ergebnissen führt. Diejenigen, die in beiden Situationen die einfache Gleichverteilungsheuristik anwandten, verdienten mehr als die strategischen Versuchspersonen.

Diese Ergebnisse stehen im Widerspruch zu denen von Güth und Van Damme (1998). Sie berichten, dass im Gegensatz zur vorliegenden Studie nur sehr geringe Beträge an den Dummy gegeben werden (S.42). Außerdem finden sie nicht eine einzige Ablehnung, die auf einen zu niedrigen Anteil für den Dummy zurückgeführt werden kann (S.230). Weitere Studien mit dem Drei-Personen-Ultimatum-Spiel sollten diese Differenzen untersuchen.

#### **5.2.2. Studie 4**

Studie 4 hat das Ziel, die Ergebnisse der Studie 3 zu replizieren und zu erweitern. Erstens wird überprüft, ob die gleichen Ergebnisse sich auch in einer nicht-studentischen (repräsentativeren) Stichprobe finden lassen. Die meisten Studien der experimentellen Spieltheorie wurden mit Studenten durchgeführt, was Zweifel über die Generalisierbarkeit der Ergebnisse aufwirft (eine Ausnahme bilden z.B. Güth et al., 2003).

Zweitens kann es einen Einfluss auf die Entscheidungen gehabt haben, dass die Entscheider in Studie 3 sowohl für sich als auch für den Dummy ein minimales Akzeptanzniveau angeben mussten. Dies könnte die Versuchspersonen dazu gebracht haben, sich moralischer zu verhalten (z.B. für sich und den Dummy den gleichen Betrag fordern). Aus diesem Grund wurde in dieser Studie das gleiche Design wie in Studie 1 gewählt (eine Variation des Drei-Personen-Bestrafungs-Spiels, s. o.). Ein bestimmter Geldbetrag musste zwischen zwei Personen (Aufteiler, Empfänger) aufgeteilt werden. Nachdem der Aufteiler den Aufteilungsvorschlag abgegeben hatte, wurde eine dritte Person darüber informiert. Die dritte Person musste angeben, ob sie die Aufteilung annimmt oder ablehnt. Die Ablehnung führte – wie in Studie 1 – dazu, dass alle drei leer ausgingen, und die

Annahme dazu, dass die Verteilung umgesetzt wurde, und die dritte Person unabhängig davon einen fixen Geldbetrag erhielt.

In diesem Versuchsaufbau macht der Aufteiler also keinen Vorschlag für die dritte Person. Deshalb ist sie auch nicht direkt von der Verteilung betroffen. Aus diesem Grund könnte die soziale Distanz zu den anderen beiden Spielern erhöht sein und die dritten Personen könnten sich sagen, dass es nicht „ihre Sache“ ist, in die Verteilung einzugreifen. Dass man trotzdem erwarten kann, dass sie eingreifen, zeigen die Ergebnisse der Studie 1. Ein weiterer Aspekt dieser Versuchsanordnung ist, dass die Kosten für altruistische Bestrafung konstant sind. In der vorherigen Studie variierten die Kosten mit der Höhe des Angebots an den Entscheider. Je höher dies war, desto kostspieliger war eine Ablehnung. Die konstanten Kosten könnten zu einer Verringerung der Bestrafung beitragen, denn für kleinere Ungleichheiten entstehen die gleichen Bestrafungskosten wie für große.

Ein weiteres Ziel dieser Studie war, mehr über die Prozesse zu erfahren, die Aufteiler dazu bringen, eine bestimmte Aufteilung vorzunehmen. Die folgenden Fragen sollten beantwortet werden:

1. Inwieweit antizipieren die Aufteiler, dass Dritte unfaire Verteilungen bestrafen? In Studie 3 wurde diese Frage eher indirekt beantwortet, indem die Angebote im Drei-Personen-Ultimatum-Spiel mit denen im Diktator-Spiel verglichen wurden. Aus diesem Grund wurden die Aufteiler diesmal nach ihrer Einschätzung gefragt, wie viel eine durchschnittliche dritte Person für den Empfänger mindestens verlangen würde.
2. Wie treffen Probanden ihre Entscheidung in der Rolle des Aufteilers? Wie oben bereits beschrieben wurde, kann man das Problem entweder als ethische Entscheidung ansehen und sich entsprechend verhalten (in diesem Fall: gleich aufteilen). Auf der anderen Seite kann man die Entscheidung aber auch strategisch angehen. Nur stellt sich dann die Frage: Wie hoch sollte man den Betrag für den Empfänger wählen? Wie stark ist das



Verhalten der Aufteilenden von ihren Einschätzungen über das Bestrafungsverhalten der dritten Person beeinflusst?

Außerdem wurde genau wie in Studie 3 die gewinnmaximierende Strategie untersucht.

### *Methode*

Die Probanden dieser Studie waren 179 Zugreisende im nördlichen Teil der Niederlande. Mit Hilfe einer quotierten Stichprobenziehung wurde sichergestellt, dass in der Stichprobe 49.2% der Versuchsteilnehmer männlich und 50.8% weiblich waren. Das Alter der Versuchspersonen lag zwischen 16 und 82 Jahren, das Durchschnittsalter in der Stichprobe betrug 41.4 Jahre ( $SD = 17.84$ ). Ungefähr 80% der Versuchsteilnehmer reisten in der 2. Klasse, ca. 20% in der ersten. Vermutlich kann man nicht sagen, dass diese Stichprobe repräsentativ für die niederländische Bevölkerung ist, aber es wurde darauf geachtet, dass sie bezüglich der Merkmale Geschlecht, Alter und sozialer Hintergrund ein breites Spektrum widerspiegelt.

Den Versuchsteilnehmern wurden beide Spiele (Diktatorspiel / Drei-Personen-Bestrafungs-Spiel) erläutert, wobei die Reihenfolge variiert wurde. Wie in Studie 3 wurde den Versuchspersonen erklärt, dass nicht für alle von ihnen die Entscheidungen auszahlungsrelevant sein würden. Ein Teil von ihnen würde über eine Lotterie ausgewählt und dann ihrer und der Entscheidung des zugelosten Partners entsprechend bezahlt.

In beiden Spielen konnten je 100 Geldeinheiten aufgeteilt werden (eine Geldeinheit entsprach ca. 0.40 Euro). Die Versuchspersonen wurden gefragt, wie viel Geld sie dem Empfänger geben würden, wenn sie in der Rolle des Aufteilenden im Drei-Personen-Bestrafungs-Spiel wären und wie viel sie mindestens als Anteil für den Empfänger vom Aufteilenden erwarten würden, wenn sie in der Rolle der dritten Person wären. Zusätzlich sollten die Versuchspersonen schätzen, was das minimale Angebot an den Empfänger ist, das eine durchschnittliche dritte Person gerade noch akzeptiert. Bei allen Fragen wurden

wieder keine Kategorien vorgegeben, so dass die Versuchsteilnehmer bei der Beantwortung völlig frei waren.

Auch hier wurden wieder mehrere Verständnisfragen gestellt, um sicher zu sein, dass die Probanden die Logik der Spiele verstanden hatten. Nach der Auswertung der Kontrollfragen mussten 13 Personen ausgeschlossen werden, so dass 166 Personen in die Analyse eingingen.

### *Ergebnisse*

Zunächst wurde die Verteilung der Angebote im Drei-Personen-Bestrafungsspiel analysiert: Im Durchschnitt gaben die Versuchspersonen 44.5 Geldeinheiten (SD = 11.08) an den Empfänger und behielten 55.5 Geldeinheiten. Eine große Mehrheit (75.2%) der Versuchspersonen teilte das Geld zu gleichen Teilen auf, während 24.8% mehr für sich verlangten als 50%.

Die Angebote an den Empfänger im Diktator-Spiel waren signifikant niedriger als die Angebote an den Empfänger im Drei-Personen-Bestrafungsspiel ( $t = 4.408$ ;  $p < .01$ ). Im Durchschnitt boten die Aufteilenden dem Empfänger 39.80 Geldeinheiten an (SD = 16.71) und behielten 60.20 Geldeinheiten für sich. Von allen Teilnehmern entschieden sich 66.2% für eine Gleichverteilung des Geldes, 33.8% behielten mehr als die Hälfte für sich.

Wie hingen die Angebote im Drei-Personen-Bestrafungsspiel mit dem Verhalten im Diktator-Spiel zusammen? Um diese Frage zu beantworten, wurde die Differenz zwischen dem Beitrag im Drei-Personen-Bestrafungsspiel und im Diktatorspiel für jede Person berechnet und mit dem Verhalten im Diktator-Spiel in Bezug gesetzt. Es zeigte sich, dass die Differenz umso höher war, je niedriger die Angebote im Diktator-Spiel waren ( $r = -.76$ ;  $p < .01$ ). Zur Verdeutlichung: Von den Personen, die im Diktator-Spiel 50/50 aufgeteilt hatten, gaben 97.7% auch im Drei-Personen-Bestrafungsspiel 50 Geldeinheiten an den Empfänger ohne Veto ab (hier gab es also keine signifikante Differenz;  $t = 1.550$ ; n. s.). Im Gegensatz dazu gaben die

10 Personen, die im Diktator-Spiel nichts abgegeben hatten, im Drei-Personen-Bestrafungsspiel im Durchschnitt 40 Geldeinheiten ab (hoch signifikante Differenz;  $t = 7.236$ ;  $p < .01$ ).

Es scheint sehr plausibel, dass diese Unterschiede von der Voraussetzung beeinflusst wurden, dass einige dritte Personen allzu ungleiche Aufteilungen im Drei-Personen-Bestrafungsspiel ablehnen könnten. Genau das ließ sich zeigen, denn die Versuchsteilnehmer gingen davon aus, dass die dritte Person im Durchschnitt mindestens 38.44 Geldeinheiten ( $SD = 13.51$ ) für den Empfänger erwarten würde.

Diese Schätzung war erstaunlich genau und unterschied sich nicht signifikant ( $t = .624$ ; n. s.) vom tatsächlich geforderten minimalen Akzeptanzbetrag von 39.09 ( $SD = 10.58$ ). Die durchschnittliche Schätzung weist aber eine starke Streuung auf, da die Einschätzungen des minimalen Akzeptanzlevels durch die Versuchspersonen Werte zwischen 0 und 50 Geldeinheiten annahm ( $SD = 13.51$ ).

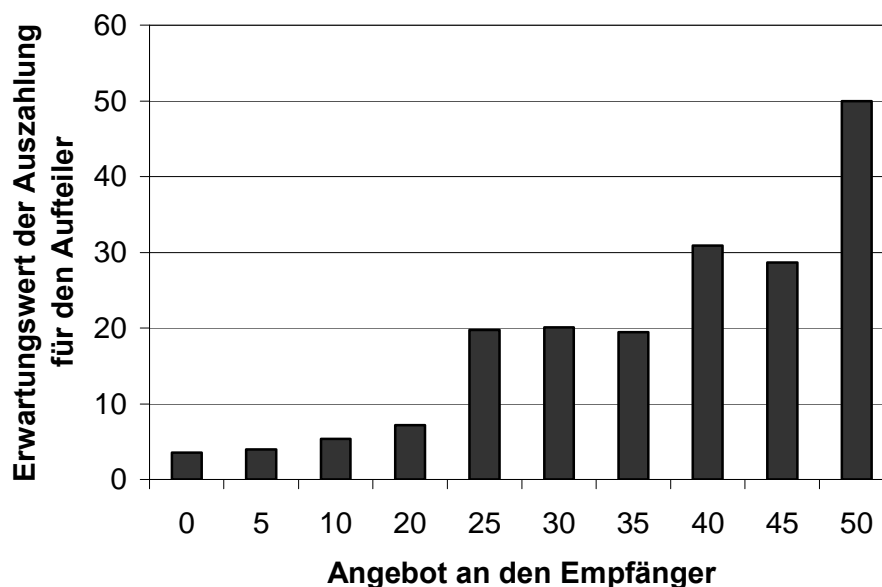
Weitere Analysen ergaben, dass die Schätzungen der Versuchsteilnehmer eng mit ihrem eigenen minimalen Akzeptanzbetrag zusammen hingen ( $r = .57$ ;  $p < .01$ ). Die Probanden scheinen also ihr eigenes Verhalten als Ausgangspunkt für ihre Schätzung des Durchschnittsverhaltens der anderen Teilnehmer zu nehmen.

Natürlich kann man einwenden, dass diese Einschätzungen lediglich „cheap talk“ waren, da keine (monetären oder anderen) Anreize gegeben wurden, die Einschätzung möglichst präzise vorzunehmen. Allerdings ist es für eine erfolgreiche Strategie im Drei-Personen-Bestrafungsspiel erforderlich, dass man die anderen valide einschätzt und dann sein Angebot auf den (geschätzten) minimalen Akzeptanzbetrag der dritten Person abstimmt. Wenn das der Fall ist, sollten die Angebote der Aufteilenden mit ihren Einschätzungen der minimalen Akzeptanzbeträge (der dritten Personen) positiv korreliert sein.

Um diese Frage zu klären, wurden aus der Stichprobe anhand des Verhaltens im Diktatorspiel zwei Gruppen gebildet: 1. Diejenigen, die 50/50 geteilt hatten, und 2. diejenigen, die mehr für sich behalten

hatten als 50%. Für diejenigen, die 50/50 geteilt hatten, wurde kein Zusammenhang erwartet, da sie offensichtlich nicht strategisch vorgehen, sondern einer einfachen Gleichverteilungsnorm folgen. Entsprechend zeigte sich in dieser Teilstichprobe auch kein Zusammenhang ( $r = .16$ ; n. s.). Für diejenigen, die im Diktator-Spiel nicht gleich geteilt hatten, die also strategisch an die Aufteilung herangingen, wurde ein Zusammenhang gefunden ( $r = .37$ ;  $p < .01$ ). Die Höhe ihrer Angebote hing signifikant mit ihrer Schätzung des minimalen Akzeptanzbetrages der dritten Personen zusammen.

Zuletzt soll wieder die Frage geklärt werden, was die gewinnmaximierende Strategie in dieser Situation war, wenn man das Ablehnungsverhalten der dritten Personen berücksichtigt. Offensichtlich hängt die Antwort mit dem Betrag zusammen, den die dritten Personen als minimalen Akzeptanzbetrag angaben, unterhalb dessen sie Angebote an den Empfänger ablehnen würden. Dieser Betrag lag bei durchschnittlich 39.06 (SD = 13.7) Geldeinheiten. Ungefähr die Hälfte der dritten Personen (46.5%) gab an, Angebote von unter 50% des aufzuteilenden Betrages an den Empfänger abzulehnen. Nur eine Minderheit von 9.6% verlangte für den Empfänger weniger als 25%. 3.6% der dritten Personen gaben an, auch Angebote von Null Geldeinheiten an den Empfänger nicht ablehnen zu wollen. Mit Hilfe dieser minimalen Akzeptanzbeträge kann nun wiederum die Wahrscheinlichkeit einer Ablehnung und damit der Erwartungswert eines Aufteilungsangebotes für den Aufteilenden ausgerechnet werden. Es zeigte sich, dass der Betrag, den der Aufteilende für sich beansprucht, und der Erwartungswert der entsprechenden Aufteilung stark negativ korreliert sind ( $r = -.97$ ;  $p < .01$ ). Abbildung 10 bestätigt auf deutliche Weise die Ergebnisse der Studie 3. Auch hier ist die 50/50-Aufteilung wieder die gewinnmaximierende Strategie, denn sie hatte eine 100%-Wahrscheinlichkeit, von den Entscheidern akzeptiert zu werden, und führte so zu einem Erwartungswert von 50 Geldeinheiten.



**Abbildung 10: Erwartungswert der Auszahlung für den Aufteilenden in Abhängigkeit vom Angebot an den Empfänger im Drei-Personen-Bestrafungs-Spiel**

Im Gegensatz dazu erbrachte eine Aufteilung 100/0 bei einer Akzeptanzwahrscheinlichkeit von 3.6% lediglich einen Erwartungswert von 3.6 Geldeinheiten.

### *Diskussion*

Die Studie 4 hatte unter anderem das Ziel, die Ergebnisse der Studie 3 zu replizieren. Dieses Ziel wurde erreicht, aber die Ergebnisse von Studie 4 gehen darüber hinaus. Es konnte erneut gezeigt werden, dass die Angebote der Aufteilenden an den Empfänger im Drei-Personen-Bestrafungs-Spiel höher waren als im Diktator-Spiel. Die Aufteiler antizipieren also, wie in Studie 3, die altruistische Bestrafung durch ihre Spielpartner. In Studie 3 konnte diese Schlussfolgerung lediglich indirekt aus dem Verhalten der Versuchspersonen abgeleitet werden. In Studie 4 wurden die Versuchspersonen daher zusätzlich nach ihrer Einschätzung der minimalen Beträge gefragt, unterhalb derer eine durchschnittliche dritte Person wohl ablehnen würde. Diese Schätzung war recht präzise und die strategisch vorgehenden Aufteiler orientierten sich an dieser Schätzung, wenn sie den

Betrag aufteilten: Je höher sie den minimalen Akzeptanzbetrag der dritten Person einschätzten, umso mehr Geld gaben sie an den Empfänger ab (einen Zusammenhang zwischen Bestrafungserwartung und Transfer finden auch Bernhard, Fehr et al., 2006).

### **5.2.3. Gesamtdiskussion der Studien 3 und 4**

In den Studien 3 und 4 konnten die Ergebnisse vieler anderer Studien repliziert werden: Menschen sind bereit, eigene (materielle) Ressourcen aufzuwenden, um unfaire Artgenossen zu bestrafen. Dieses Verhalten zeigt sich sowohl, wenn sie selbst betroffen sind, aber eben auch, wenn sie sich für eine andere unfair behandelte Person einsetzen.

In Studie 3 zeigte sich, genau wie bei Fehr und Fischbacher (2004b), die allerdings ein anderes Spiel-Design verwenden, dass die Bereitschaft zu Third-Party-Punishment zwar geringer ist, dass es aber trotzdem zu einer nicht unerheblichen Bestrafung durch Dritte kommt. In Studie 3 waren 51.4% aller Teilnehmer nicht bereit, eine Aufteilung zu akzeptieren, die für den Dummy weniger als  $1/3$  der aufzuteilenden Gesamtsumme vorsah. Im Gegensatz dazu gaben lediglich 12.2% der VPn an, dass sie jeden Betrag (also auch Null) für den Dummy akzeptieren würden. In Studie 4 zeigte sich ein ähnliches Bild: Fast die Hälfte der Versuchspersonen akzeptierte nur eine Gleichverteilung.

Beim Vergleich der Ergebnisse der Studien 3 und 4 muss berücksichtigt werden, dass sie sich sowohl durch das verwendete experimentelle Spiel als auch durch die Stichprobe unterscheiden (deutsche Studenten vs. niederländische Zugreisende). Umso bemerkenswerter ist, dass der Wunsch nach Gleichverteilung und die Bereitschaft zu bestrafen in beiden Untersuchungen ähnlich hoch ist.

Hinsichtlich der Widersprüche der Ergebnisse zu denen von Güth und van Damme (1998) muss zunächst gesagt werden, dass sowohl Fehr und Fischbacher (2004b) als auch Fetchenhauer und Huang

(2004) ein recht hohes Maß an Bestrafung finden und dies auf Gerechtigkeitsstreben der Beteiligten zurückführen. Mit dem Drei-Personen-Ultimatum-Spiel bzw. Drei-Personen-Bestrafungs-Spiel sind noch nicht so viele Studien durchgeführt worden wie z.B. mit dem Ultimatum-Spiel. Auch dort gibt es immer wieder unterschiedliche Ergebnisse, die auf verschiedene Ursachen zurückgeführt werden (Design, Stichprobe etc.). Für das Drei-Personen-Ultimatum-Spiel liegen meines Erachtens noch nicht genügend Studien vor, um sagen zu können, wie das Verhalten der Versuchspersonen im Durchschnitt aussieht. Weitere Studien, die insbesondere Veränderungen des Designs berücksichtigen, sollten diese Frage empirisch klären.

Die Ergebnisse dieser beiden Studien ähneln denen von Fehr und Fischbacher (2004b) in Bezug auf die weit verbreitete Bereitschaft zu Second- und Third-Party-Punishment. Allerdings unterscheiden sie sich in einem entscheidenden Punkt: In der Studie von Fehr und Fischbacher kann ein Aufteiler im Drei-Personen-Bestrafungs-Spiel seinen Gewinn maximieren, indem er dem Empfänger weniger als den gleichen Anteil anbietet (die Zusatz-Gewinne für diese unfaire Strategie sind allerdings nicht sehr hoch). In den hier durchgeführten Studien 3 und 4 war es so, dass die Gleichverteilung die gewinnmaximierende Strategie ist. Dieses abweichende Ergebnis kann allerdings auf einen kleinen, aber wichtigen Unterschied im Design zurückgeführt werden. Die Studie 4 unterschied sich insofern von der von Fehr und Fischbacher, als dort Bestrafungspunkte einzeln zugewiesen werden konnten, die zu einem Abzug im Verhältnis 1:3 (s. o.) beim Aufteiler führten. Die Bestrafung war also dosierbar. In Studie 4 war nur eine Annahme oder Ablehnung der gesamten Verteilung möglich. Bei Fehr und Fischbacher war es also möglich, die Aufteiler auch in sehr kleinem Umfang zu bestrafen. Vielleicht nutzten einige Versuchspersonen die Möglichkeit, ihre Kritik am Aufteiler auszudrücken ohne auf allzu viele Ressourcen verzichten zu müssen. Beide experimentellen Designs haben ihre Existenzberechtigung. Es sind

Situationen vorstellbar, in denen nur nach dem Alles-oder-Nichts-Prinzip verhandelt werden kann, es also keine Abstufungen gibt. Auf der anderen Seite mag es aber auch Situationen geben, in denen man weiter verhandelt, allerdings über ein geringeres Auftragsvolumen als man es zunächst vorgesehen hatte.

Die Hauptfrage der Studien 3 und 4 war: Antizipieren Menschen die Bereitschaft zu altruistischer Bestrafung durch ihre Mitmenschen? Die Antwort ist ein klares „Ja“! In beiden Studien waren die Angebote für den Dummy bzw. Empfänger (in den jeweiligen Drei-Personen-Spielen) höher als die Angebote für den ebenfalls machtlosen Empfänger im Diktator-Spiel. Hieraus lässt sich ableiten, dass die Aufteilenden vorhersahen, dass die Entscheider bzw. dritten Personen sie für (sehr) ungleiche Verteilungen bestrafen würden.

Außerdem scheinen die Versuchspersonen ein Gespür dafür zu haben, dass die Entscheider eher bereit sind zu bestrafen, wenn sie selbst ungerecht behandelt wurden (SPP), als wenn sie nicht selbst von der ungerechten Aufteilung betroffen sind (TPP). Dies lässt sich daraus ableiten, dass die Angebote für den Entscheider in Studie 3 höher waren als für den Dummy. Solche strategischen Erwägungen hängen mit dem Grad intrinsischer Fairness der Versuchspersonen zusammen – gemessen darüber, welchen Anteil eine Versuchsperson in der Rolle des Aufteilenden in einem Diktatorspiel an den Empfänger gibt. Da es im Diktator-Spiel keinen externen, materiellen Anreiz gibt, sich fair zu verhalten (i. e. 50/50-Aufteilung), wurde in beiden Studien von den Versuchspersonen, die dies dennoch taten, angenommen, dass sie sich auch in den jeweiligen Drei-Personen-Spielen fair verhalten würden. Im Gegensatz dazu wurde von denjenigen, die im Diktator-Spiel nicht gleich aufteilten, angenommen, dass sie die Drei-Personen-Spiele strategisch angehen würden. Beide Studien stützen diese Hypothese: In Studie 3 gaben nur diejenigen, welche strategisch aufteilten, dem Dummy weniger als dem Entscheider. In beiden Studien gaben strategische Personen dem Empfänger im



Diktator-Spiel weniger als dem Entscheider bzw. der dritten Person im jeweiligen Drei-Personen-Spiel. Im Gegensatz dazu teilten die Aufteiler, die als fairnessmotiviert klassifiziert worden waren, das Geld in beiden Studien gleichmäßig auf alle Beteiligten auf.

Das bemerkenswerteste Ergebnis ist allerdings, dass die faire Aufteilung auch das Ergebnis des Aufteilenden maximierte, ein Befund, der sich mittlerweile in verschiedenen Studien mit verschiedenen Spielen gezeigt hat. Zumindest für westliche Gesellschaften gilt, dass eine faire Aufteilung des Geldbetrages im Ultimatum-Spiel zum höchsten Gewinn des Aufteilers und des Empfängers führt (Fehr & Fischbacher, 2004b; Güth et al., 2003; van't Wout et al., 2006; Prasnika & Roth, 1992; Harrison & McCabe, 1996; Bolton & Zwick, 1995; Kravitz & Gunto, 1992). Begründet wird das damit, dass die Ablehnungswahrscheinlichkeit steigt, je weiter sich ein Angebot von der fairen Aufteilung fortbewegt, somit sinkt dann der (Erwartungs-) Ertrag. Gürer, Irlenbusch und Rockenbach (2006) konnten für ein Öffentliches-Gut-Spiel mit Bestrafung nachweisen, dass diejenigen, die voll kooperierten und zum öffentlichen Gut beitrugen, mehr verdienten als die Trittbrettfahrer. Bei Third-Party-Punishment ist das Bild nicht ganz so einheitlich. In dem von Fehr und Fischbacher (2004b) geschilderten Versuch lohnte sich eigennütziges Verhalten der Aufteilenden, da die Abzüge durch die Bestrafung nicht die gesamten Mehreinnahmen aus der ungleichen Verteilung vernichteten. Hierzu ist allerdings anzumerken, dass bei einem veränderten Design mit härterer Wirkung der Bestrafung vermutlich auch dieser kleine Gewinn verloren gehen würde, so dass sich das unfaire Aufteilen dann auch bei Bestrafung durch unbeteiligte Dritte nicht mehr lohnen würde.

In den Studien 3 und 4 sind die höheren Erträge für faire Aufteiler nicht darauf zurückzuführen, dass die strategischen Aufteiler die Bestrafungsneigung nicht berücksichtigen. Das tun sie, was die unterschiedlichen Angebote für Entscheider bzw. dritte Person und

Dummy sowie die Zusammenhänge der Angebote mit den Schätzungen der minimalen Akzeptanzbeträge zeigen. Allerdings unterschätzen sie diese Bereitschaft stark. Ein Grund dafür ist der starke „Falsche Consensus-Effekt“, dem die Versuchspersonen erliegen. Sie gehen davon aus, dass die anderen sich so verhalten wie sie selbst, wenn sie den minimalen Akzeptanzbetrag der anderen schätzen sollen. Es muss zudem berücksichtigt werden, dass ihr eigener minimaler Akzeptanzbetrag stark mit den eigenen Angeboten in der Rolle des Aufteilenden korreliert ( $r = .57$ ;  $p < .01$ ). Diese egozentrische Annahme lässt sie davon ausgehen, die anderen würden ähnlich niedrige Angebote akzeptieren, wie sie es selbst täten.

Hier wird klar, dass es einen Unterschied macht, ob ein Spiel nur einmal gespielt wird oder ob es wiederholt mit den gleichen Partnern gespielt wird. Es ist anzunehmen, dass die „Strategen“ aus Studie 3 und 4 mit der Zeit die Bestrafungsneigung weniger unterschätzen würden. Dann könnten sie ihre Strategie besser anpassen und ihre Erträge würden vermutlich steigen.

Beide Studien beruhen auf der Annahme, dass die Bestrafung nur dann eine positive Wirkung für die Gesellschaft haben kann, wenn sie von den Akteuren antizipiert wird. Da die Strafe mit Kosten verbunden ist, sinkt der Gesamtertrag aller Beteiligten. Muss allerdings aufgrund verbreiteter Normeinhaltung und aufgrund der abschreckenden Wirkung nicht mehr sehr häufig bestraft werden, verringern sich diese Verluste für die Gesellschaft.

## 6. Schlussfolgerungen und Ausblick

Zu Beginn dieser Arbeit wurde die Frage aufgeworfen, wie Normen der Fairness oder der Kooperation in einer Gesellschaft durchgesetzt werden. Die Antwort dieser Arbeit ist, dass Normbrecher bestraft werden, und zwar entweder durch die Person, die Schaden genommen hat, oder durch eine dritte Partei. Viele empirische Studien, über die in dieser Arbeit berichtet wurde, stützen diese Sichtweise. Die vier eigenen Studien erweitern diese empirische Basis in verschiedener Hinsicht: Die Studien 1 und 2 stellen zum ersten Mal die Verbindung zwischen einem Persönlichkeitsmerkmal (Ungerechtigkeitssensibilität), den negativen Emotionen in ungerechten Situationen und der altruistischen Bestrafung her. Die Studien 3 und 4 untersuchen explizit die Bestrafungserwartungen der Versuchspersonen und deren Reaktion darauf.

Die meisten Versuchspersonen in den einzelnen Untersuchungen hatten offensichtlich relativ genaue Vorstellungen davon, wie das Geld in den experimentellen Spielen aufgeteilt werden sollte. Dies ist ein starker Hinweis auf eine allgemeine Norm über die Verteilung von Erträgen in solchen Situationen. Dies wird in den vorgestellten Studien an vier Punkten deutlich:

1. Wenn die Versuchspersonen selbst aufteilen müssen, dominiert die Gleichverteilung.
2. Wenn die Versuchspersonen bestrafen können, werden Abweichungen von der Gleichverteilung bestraft.
3. In den Kommentaren machen die Versuchspersonen deutlich, wie das Geld aufgeteilt werden sollte und was sie von Menschen halten, die nicht so aufteilen.
4. Selbst wenn sie nicht bestrafen, geben manche Versuchspersonen an bereit zu sein, einer ungerecht behandelten Person Geld zu geben.

Für viele Menschen scheint es also eine bindende Verteilungsnorm zu geben, an die sie sich unabhängig von Eigennutzerwägungen halten.

Alle vier eigenen Studien belegen zudem zweifelsfrei die Bereitschaft, unbekannte andere Personen für eine ungerechte Aufteilung zu bestrafen. Bestraft wird sogar dann, wenn dies mit eigenen Kosten verbunden ist und die anderen Versuchspersonen völlig anonym sind. Außerdem wird nicht nur eigene erlebte Ungerechtigkeit geahndet, sondern auch unfaires Verhalten gegenüber anderen. Für drei studentische und eine nicht-studentische Stichprobe zeigten sich gleiche Bestrafungsmuster. Die Resultate sind im Einklang mit einer Vielzahl anderer Studien zu verschiedenen experimentellen Spielen, die altruistische Bestrafung nachweisen konnten.

*Tunica propior palliost* – Das Hemd ist mir näher als der Rock. Bemerkenswert, allerdings nicht sonderlich überraschend ist, dass die Versuchspersonen mehr bestrafen, wenn sie selbst betroffen sind. Wenn sie einen Normbruch lediglich beobachten und als dritte Partei eingreifen können, ist ihre Bestrafungsbereitschaft geringer. Obwohl für beide Arten der Bestrafung die gleichen Mechanismen vermutet werden, ist die Neigung, in eigener Sache zu bestrafen, also höher, als wenn für andere bestraft wird.

Wird mit der Bestrafung durch andere überhaupt gerechnet und wirkt sich das auf das Verhalten aus? Ja, wie auch schon in anderen Studien wird hier in Ultimatum-Spielen deutlich mehr an den Entscheider gezahlt als an den Empfänger in einem Diktator-Spiel. Der Aufteiler fürchtet offensichtlich die Bestrafung / Ablehnung durch den Entscheider im Ultimatum-Spiel und gibt daher mehr.

Dieses strategische Verhalten zeigt sich auch im Vergleich des Diktator-Spiels mit den Drei-Personen-Spielen (Drei-Personen-Ultimatum-Spiel, Drei-Personen-Bestrafungs-Spiel). Ein Teil der Aufteiler gibt der machtlosen Person in den Drei-Personen-Spielen mehr Geld als im Diktator-Spiel, weil sie die Bestrafung antizipieren. Das bedeutet, dass auch Third-Party-Punishment erwartet wird.

Ist es gewinnmaximierend sich strategisch zu verhalten oder zahlt sich Fairness am Ende aus? Die Frage, ob Fairness sich auszahlt, kann nicht abschließend beantwortet werden, allerdings deuten die Ergebnisse darauf hin. Strategisches Verhalten in sozialen Dilemmasituationen wäre vermutlich gewinnmaximierend, wenn die Strategie der Aufteiler auf einer korrekten Einschätzung der Bestrafungsbereitschaft basieren würde. Diese Bereitschaft wird aber unterschätzt und somit wird zu wenig angeboten. Zu niedrige Angebote werden abgelehnt, so dass dann der Ertrag völlig ausbleibt. Empirisch sieht es häufig so aus, dass in experimentellen Spielen strategisches Verhalten auf der Basis korrekter Einschätzungen zum gleichen Ergebnis führt wie faires Verhalten. Intrinsisch fairnessmotivierte Individuen haben noch weitere Vorteile gegenüber strategischen Individuen: Sind Fairnessnormen internalisiert, dürfte die Entscheidung für eine bestimmte Verteilung mehr oder weniger automatisch und schnell erfolgen, bei der strategischen Entscheidung hingegen ist deutlich mehr kognitiver Aufwand nötig. Man muss seinen Interaktionspartner einschätzen und das Angebot darauf abstimmen, und wenn man sich verschätzt hat, ist die Gefahr der Ablehnung hoch. Die Schnelligkeit und der hohe Erwartungswert fairer Entscheidungen, die zu nahezu einhundert Prozent angenommen werden, sind klare Vorteile für ein faires Individuum.

Die Bestrafungsbereitschaft hängt zumindest für einige Menschen von den Kosten bzw. der Wirkung der Bestrafung ab. Die Tatsache, dass es auch außerhalb des Labors zu Bestrafung durch Dritte kommt und wir in weiten Bereichen einen hohen Grad an Normeinhaltung beobachten können, legt die Vermutung nahe, dass die Kosten der Bestrafung durch Dritte in der realen Welt nicht prohibitiv hoch sind und die Bilanz aus Kosten und Nutzen der Bestrafung positiv ausfällt. Eine andere Erklärung ist, dass es genügend Menschen gibt, die (relativ) unabhängig von Kosten-Nutzen-Überlegungen, aufgrund deontischer Gerechtigkeitsvorstellungen bestrafen.

Um diese Personen zu identifizieren, sollte neben den unbestritten wichtigen situativen Faktoren (Kosten, Kommunikation usw.) der Einfluss der Persönlichkeit bei der Aufteilung und der Bestrafung in experimentellen Spielen zukünftig noch stärkere Beachtung finden. Die Ungerechtigkeitssensibilität (US) nach Schmitt et al. (1997) hat sich bereits mehrfach als guter Prädiktor für prosoziales Verhalten sowohl in experimentellen Spielen als auch im Alltagskontext erwiesen.

Grundsätzlich sollte man sich bei der Untersuchung des Zusammenhangs von Persönlichkeitsvariablen und Entscheidungen in experimentellen Spielen klar machen, für welche Positionen in einzelnen experimentellen Spielen eindeutige Vorhersagen aufgrund der Persönlichkeit möglich sind. Spannend sind insbesondere die Positionen in Spielen, auf denen die Verfolgung des Fairness- oder des Eigennutzmotivs zu entgegengesetztem Verhalten führen können. Für die Position des Aufteilers im Ultimatum-Spiel und im Drei-Personen-Bestrafungs-Spiel (und in einigen anderen Spielen) ist dies offensichtlich nicht der Fall: Eine 50/50-Aufteilung in einem Ultimatumspiel kann von einem sehr fairen Spieler oder von einem strategischen Spieler, der seinem Gegenspieler eine starke Bestrafungsneigung unterstellt, vorgenommen werden. Zwei völlig unterschiedliche Motivationen bzw. Spielerpersönlichkeiten führen zum gleichen beobachtbaren Verhalten. Anders sieht es für das Diktator-Spiel bzw. für die Position des Entscheiders im Ultimatum-Spiel (und in anderen Spielen) aus. Hier ergibt sich ein unterschiedliches Verhalten abhängig davon, ob ein Spieler strategisch nach Gewinnmaximierung strebt oder ob er an Fairness interessiert ist. Die Variation im Verhalten, kann offensichtlich mit der Persönlichkeit der Spieler erklärt werden.

Wenn es eine allgemein bekannte Verteilungsnorm gibt, warum handeln dann nicht alle Personen in den Experimenten gleich? Die Kenntnis der Norm an sich stellt noch keinen genügend großen Anreiz dar, sich auch normkonform zu verhalten. Damit eine Norm

verhaltenswirksam wird, muss sie entweder internalisiert sein oder ihre Beachtung muss permanent durch die Gesellschaft überwacht werden. Im Fall der Norminternalisierung führen moralische Emotionen zu normkonformem Verhalten. Wer sich angesichts der Verletzung einer Fairnessnorm schlecht fühlt (Ärger, Empörung), wird sich selbst an die Norm halten und andere für einen Normbruch bestrafen. Menschen, die eine Norm nicht internalisiert haben, werden sich dann an die Norm halten, wenn sie die Wahrscheinlichkeit bestraft zu werden hoch einschätzen; sie werden sich aber nicht an die Norm halten, wenn ein Normbruch sich für sie lohnt. Manche Menschen verhalten sich z.B. nur in der Öffentlichkeit normkonform, nicht aber wenn sie sich unbeobachtet fühlen. Andere wiederum halten sich immer an bestimmte Normen, unabhängig davon, ob sie anonym bleiben oder nicht. Für die erste Gruppe von Menschen kann man nicht sagen, ihr „öffentliches“ oder ihr „privates“ Verhalten sei ihr wirkliches Verhalten, sondern nur, ihr Verhalten sei situationsabhängig.

Wie die Studien 1 und 2 zeigen, sind Außenstehende offensichtlich in der Lage, die negativen Emotionen aus den Kommentaren der Versuchspersonen zu entnehmen. Erkennt man negative moralische Emotionen bei einem Interaktionspartner, kann man sein Bestrafungsverhalten einschätzen. Wer stark negativ auf Ungerechtigkeit reagiert und seine Emotionen zeigt, ist entsprechend als eine Person zu identifizieren, die altruistisch bestraft. Im Sinne des Commitment-Modells von Frank kann dies dazu führen, dass diese Personen in Zukunft von fairen Interaktionspartnern bevorzugt werden. Dieser Punkt und die Tatsache, dass sich in verschiedenen Studien mit experimentellen Spielen die Gleichverteilung als gewinnmaximierend herausstellte, ist eine starke Unterstützung für das Commitment-Modell.

Im Vordergrund zukünftiger Forschungsvorhaben sollte die Identifizierung möglichst vieler Situations- und Persönlichkeitsvariablen

stehen, die einen Einfluss auf Aufteilung und Bestrafung in experimentellen Spielen haben. Z.B. sollte empirisch festgestellt werden, ab welchem Kosten-Wirkungs-Verhältnis der Bestrafung sich eine Abweichung von der Gleichverteilung nicht mehr lohnt.

Die in dieser Arbeit versuchte Zusammenführung der Theorien und der empirischen Methoden verschiedener Disziplinen für die Erforschung menschlicher Verhaltensweisen in sozialen Dilemmasituationen stellt sich als fruchtbar heraus. Die Verfolgung monokausaler Argumentationsstränge, die jeweils nur isolierten Facetten des komplexen Menschen gerecht werden, z.B. seinem Eigennutzstreben, müssen zugunsten von Ansätzen aufgegeben werden, die auf einer breiteren (psychologischen) Basis stehen.



## 7. Literatur

- Abbink, K., Irlenbusch, B. & Renner, E. (2000). The moonlighting game - An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, 42(2), 265-277.
- Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, 2(S 267), 299.
- Ajzen, I. (1988). *Attitudes, Personality and Behavior*. Chicago: The Dorsey Press.
- Andreoni, J. (1995). Cooperation in Public-Goods Experiments - Kindness or Confusion. *American Economic Review*, 85(4), 891-904.
- Austin, W. & Walster, E. (1974). Reactions to Confirmations and Disconfirmations of Expectancies of Equity and Inequity. *Journal of Personality and Social Psychology*, 30(2), 208-216.
- Axelrod, R. & Dion, D. (1988). The Further Evolution of Cooperation. *Science*, 242(4884), 1385-1390.
- Axelrod, R. & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, 211(4489), 1390-1396.
- Bahry, D. L. & Wilson, R. K. (2006). Confusion or fairness in the field? Rejections in the ultimatum game under the strategy method. *Journal of Economic Behavior & Organization*, 60(1), 37-54.
- Baker, K. (1974). Experimental Analysis of Third-Party Justice Behavior. *Journal of Personality and Social Psychology*, 30(2), 307-316.
- Baron, R. M. & Kenny, D. A. (1986). The Moderator Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H. & Strongman, J. A. (1999). Moral hypocrisy: Appearing Moral to Oneself Without Being So. *Journal of Personality and Social Psychology*, 77(3), 525-537.

- Baumert, A. (2003). *Kognitive Prozesse der Ungerechtigkeitssensibilität aus der Beobachterperspektive - Selektive Informationssuche*. Unveröffentlichte Diplomarbeit. Universität Trier.
- Bendor, J. & Mookherjee, D. (1990). Norms, Third-Party Sanctions, and Cooperation. *Journal of Law Economics & Organization*, 6(1), 33-63.
- Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, Reciprocity, and Social-History. *Games and Economic Behavior*, 10(1), 122-142.
- Bernhard, H., Fehr, E. & Fischbacher, U. (2006). Group Affiliation and Altruistic Norm Enforcement. *American Economic Review*, 96(2), 217-221.
- Bernhard, H., Fischbacher, U. & Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105), 912-915.
- Biel, A. & Thøgersen, J. (2007). Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, 28(1), 93-112.
- Bierhoff, H. W. (2002). Prosoziales Verhalten. In W. Stroebe, K. Jonas & M. Hewstone (Hrsg.), *Sozialpsychologie. Eine Einführung* (4. Aufl., S. 319-351). Berlin: Springer-Verlag.
- Bolton, G. E. & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1), 166-193.
- Bolton, G. E. & Zwick, R. (1995). Anonymity Versus Punishment in Ultimatum Bargaining. *Games and Economic Behavior*, 10(1), 95-121.
- Bosman, R., Sonnemans, J. & Zeelenberg, M. (2001). *Emotions, rejections, and cooling off in the ultimatum game*. Unpublished Work. Center for Research in Experimental Economics and Political Decision-Making (CREED). University of Amsterdam.
- Bosman, R., Sutter, M. & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology*, 26(3), 407-429.

- Bosman, R. & van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal*, 112(476), 147-169.
- Bowles, S. & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1), 17-28.
- Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6), 3531-3535.
- Boyd, R. & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *J Theor Biol*, 132(3), 337-356.
- Brandstätter, H. & Güth, W. (2002). Personality in dictator and ultimatum games. *Central European Journal of Operations Research*, 10, 191-215.
- Brandstätter, H. & Königstein, M. (2001). Personality Influences on Ultimatum Bargaining Decisions. *European Journal of Personality*, 15(1), S53-S70.
- Brandts, J. & Charness, G. (2000). Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games. *Experimental Economics*, 2(3), 227-238.
- Camerer, C. F. (2003). *Behavioral Game Theory*. New York: Russell Sage Foundation.
- Camerer, C. F. & Fehr, E. (2004). Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists. In J. Henrich, R. Boyd, S. Bowles, C. F. Camerer, E. Fehr & H. Gintis (Eds.), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (pp. 55-95). Oxford: Oxford University Press.
- Camerer, C. F. & Hogarth, R. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3), 7-42.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

- Cox, D. R. & Snell, E. J. (1970). *Analysis of Binary Data*. London: Chapman and Hall.
- Cropanzano, R., Goldman, B. & Folger, R. (2003). Deontic justice: the role of moral principles in workplace fairness. *Journal of Organizational Behavior*, 24, 1019-1024.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Son.
- Dawes, R. M. (1980). Social Dilemmas. *Annual Review of Psychology*, 31, 169-193.
- de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., et al. (2004). The Neural Basis of Altruistic Punishment. *Science*, 305(5688), 1254-1258.
- Diekmann, A. (2004). The Power of Reciprocity - Fairness, reciprocity, and stakes in variants of the dictator game. *Journal of Conflict Resolution*, 48(4), 487-505.
- Elster, J. (1990). Norms of Revenge. *Ethics*, 100(4), 862-885.
- Elster, J. (1994). Rationality, Emotions, and Social Norms. *Synthese*, 98(1), 21-49.
- Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature*, 36(1), 47-74.
- Falk, A., Fehr, E. & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, 41(1), 20-26.
- Falk, A., Fehr, E. & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73(6), 2017-2030.
- Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293-315.
- Fehr, E. (2004). Don't lose your reputation. *Nature*, 432(7016), 449-450.
- Fehr, E. & Fischbacher, U. (2002). Why social preferences matter - The impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478), C1-C33.

- Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785-791.
- Fehr, E. & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4), 185-190.
- Fehr, E. & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63-87.
- Fehr, E., Fischbacher, U. & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature-an Interdisciplinary Biosocial Perspective*, 13(1), 1-25.
- Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. & Wagner, G. G. (2002). A Nation-Wide Laboratory. *Schmollers Jahrbuch*, 122, 519-542.
- Fehr, E. & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4), 980-994.
- Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137-140.
- Fehr, E. & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3), 817-868.
- Fetchenhauer, D. (2001). Fehlende Solidarität: Zur Erklärung von Steuerhinterziehung, Schwarzarbeit und (Sozial-)Versicherungsbetrug. In H. W. Bierhoff & D. Fetchenhauer (Hrsg.), *Solidarität. Konflikt, Umwelt und Dritte Welt* (205-230). Opladen: Leske + Budrich.
- Fetchenhauer, D. & Bierhoff, H. W. (2004). Altruismus aus evolutionstheoretischer Perspektive. *Zeitschrift für Sozialpsychologie*, 35(3), 131-141.
- Fetchenhauer, D. & Dunning, D. (2005). *Do people trust too much or too little?* Unpublished manuscript.
- Fetchenhauer, D., Gresser, F. N. & Haferkamp, A. (in Druck). Zwischen Zynismus und Gutgläubigkeit –Vertrauen wir anderen zu viel oder zu wenig? In F. Rexroth et al. (Hrsg.), *Kooperation und Betrug in Beziehungssystemen*.

- Fetchenhauer, D. & Groothuis, T. (2007). *Twenty seconds are enough: identifying others' altruism by thin slices of behaviour*. Unveröffentlichtes Manuskript. Universität zu Köln.
- Fetchenhauer, D. & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences*, 36(5), 1015-1029.
- Fetchenhauer, D. & van der Vegt, G. (2001). Honesty, Trust and Economic Growth. A Cross-Cultural Comparison of Western Industrialized Countries. *Zeitschrift für Sozialpsychologie*, 32(3), 189-200.
- Fischbacher, U., Gächter, S. & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397-404.
- Fischer, L. & Wiswede, G. (2002). *Grundlagen der Sozialpsychologie* (2. überarbeitete und erweiterte Auflage). München: R. Oldenbourg Verlag.
- Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3), 347-369.
- Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. New York: W.W. Norton & Company.
- Frank, R. H. (2004). *What Price the Moral High Ground? Ethical Dilemmas in Competitive Environments* Princeton: Princeton University Press.
- Frank, R. H., Gilovich, T. & Regan, D. T. (1993). The Evolution of One-Shot Cooperation - an Experiment. *Ethology and Sociobiology*, 14(4), 247-256.
- Frijda, N. H. (1988). The Laws of Emotion. *American Psychologist*, 43(5), 349-358.
- Gaulin, S. J. C. & McBurney, D. H. (2004). *Evolutionary Psychology* (2nd ed.). New Jersey: Prentice Hall.
- Gintis, H. (2000). Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206(2), 169-179.

- Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24(3), 153-172.
- Gneezy, U. & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3), 791-810.
- Goette, L., Huffman, D. & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, 96(2), 212-216.
- Gollwitzer, M. (2004). *Eine Analyse von Racheaktionen und rachebezogenen Reaktionen unter gerechtigkeitspsychologischen Aspekten (Dissertation)*. Trier: Universität Trier.
- Gollwitzer, M., Schmitt, M., Fetchenhauer, D., Baumert, A. & Schlösser, T. (in Druck). Soziale Gerechtigkeit. In N. Goldschmidt & H. G. Nutzinger (Hrsg.), *Vom homo oeconomicus zum homo culturalis. Handlung und Verhalten in der Ökonomie*. Berlin: LIT Verlag.
- Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J. & Baer, A. (2005). Asymmetrical Effects of Justice Sensitivity Perspectives on Prosocial and Antisocial Behavior. *Social Justice Research*, 18(2), 183-201.
- Greene, J. & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517-523.
- Gürerk, O., Irlenbusch, B. & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770), 108-111.
- Güth, W., Schmittberger, R. & Schwarze, B. (1982). An Experimental-Analysis of Ultimatum Bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367-388.
- Güth, W., Sutter, M. & Schmidt, C. (2003). Fairness in the Mail and Opportunism in the Internet: A Newspaper Experiment on Ultimatum Bargaining. *German Economic Review*, 4(2), 243-265.
- Güth, W. & van Damme, E. (1998). Information, Strategic Behavior, and Fairness in Ultimatum Bargaining: An Experimental Study. *Journal of Mathematical Psychology* 42(2/3), 227-247.

- Hagen, E. H. & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, 69(3), 339-348.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.
- Hamilton, W. D. (1963). Evolution of Altruistic Behavior. *American Naturalist*, 97(896), 354-356.
- Hamilton, W. D. (1964). Genetical Evolution of Social Behaviour I. *Journal of Theoretical Biology*, 7(1), 1-16.
- Harrison, G. W. & McCabe, K. A. (1996). Expectations and fairness in a simple bargaining experiment. *International Journal of Game Theory*, 25(3), 303-327.
- Hassebrauck, M. (1984). *Emotionale Konsequenzen distributiver Ungerechtigkeit: Experimentelle Untersuchung zur Equity-Theorie*. Regensburg: S. Roderer Verlag.
- Heckathorn, D. D. (1988). Collective Sanctions and the Creation of Prisoners-Dilemma Norms. *American Journal of Sociology*, 94(3), 535-562.
- Heckathorn, D. D. (1989). Collective Action and the Second-Order Free-Rider Problem. *Rationality and Society*, 1(1), 78-100.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons.
- Henrich, J. & Boyd, R. (2001). Why people punish defectors - Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208(1), 79-89.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2), 73-78.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2005). "Economic man" in cross-cultural perspective:



- Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795-815.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767-1770.
- Hertwig, R. & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(3), 383-451.
- Hoffman, E., McCabe, K. A. & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, 25(3), 289-301.
- Hogan, R. & Emler, N. P. (1981). Retributive Justice. In M. J. Lerner & S. C. Lerner (Eds.), *The Justice Motive in Social Behavior*. New York: Plenum Press.
- Homans, G. C. (1961). *Social Behavior: Its Elementary Forms*. New York: Harcourt, Brace & World.
- Hopfensitz, A. & Reuben, E. (2005). *The Importance of Emotions for the Effectiveness of Social Punishment*. Discussion Paper. Nr.: TI 2005-075/1. University of Amsterdam and Tinbergen Institute.
- Horne, C. (2001). The enforcement of norms: Group cohesion and meta-norms. *Social Psychology Quarterly*, 64(3), 253-266.
- Horne, C. (2003). The internal enforcement of norms. *European Sociological Review*, 19(4), 335-343.
- Horne, C. (2004). Collective benefits, exchange interests, and norm enforcement. *Social Forces*, 82(3), 1037-1062.
- Horne, C. & Cutlip, A. (2002). Sanctioning costs and norm enforcement - An experimental test. *Rationality and Society*, 14(3), 285-307.
- Houser, D. & Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *American Economic Review*, 92(4), 1062-1069.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

- Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1986a). Fairness and the Assumptions of Economics. *Journal of Business*, 59(4), S285-S300.
- Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1986b). Fairness as a Constraint on Profit Seeking - Entitlements in the Market. *American Economic Review*, 76(4), 728-741.
- Kahneman, D. & Tversky, A. (1979). Prospect Theory - Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.
- Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten*. Riga.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829-832.
- Kohlberg, L. (1968/1995). Moralische Entwicklung. In W. Althof, G. Noam & F. Oser (Hrsg.), *Lawrence Kohlberg: Die Psychologie der Moralentwicklung* (S. 7-40). Frankfurt am Main: Suhrkamp Verlag.
- Kohlberg, L. (1976/1995). Moralstufen und Moralerwerb: Der kognitiv-entwicklungstheoretische Ansatz. In W. Althof, G. Noam & F. Oser (Hrsg.), *Lawrence Kohlberg: Die Psychologie der Moralentwicklung* (S. 123-174). Frankfurt am Main: Suhrkamp Verlag.
- Kohlberg, L. & Candee, D. (1984/1995). Die Beziehung zwischen moralischem Urteil und moralischem Handeln. In W. Althof, G. Noam & F. Oser (Hrsg.), *Lawrence Kohlberg: Die Psychologie der Moralentwicklung* (S. 373-493). Frankfurt am Main: Suhrkamp Verlag.
- Kravitz, D. A. & Guntto, S. (1992). Decisions and Perceptions of Recipients in Ultimatum Bargaining Games. *Journal of Socio-Economics*, 21(1), 65-84.
- Lerner, M. J. (1977). The justice motive - Some hypotheses as to Its origins and forms. *Journal of Personality*, 45(1), 1-52.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K. & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267-286.
- Luhmann, N. (2000). *Vertrauen* (4. Aufl.). Stuttgart: Lucius & Lucius Verlagsgesellschaft.

- Mackie, J. L. (1985). *Persons and values* (Vol. II). Oxford: Clarendon Press.
- Marwell, G. & Schmitt, D. R. (1968). Are "trivial" games the most interesting psychologically? *Behav Sci*, 13(2), 125-128.
- Masclet, D. (2003). Ostracism in work teams: a public good experiment. *International Journal of Manpower*, 24(7), 867-887.
- Masclet, D. & Villeval, M. C. (2006). *Punishment, Inequality and Emotions*. Discussion Paper. Nr.: 2119. Forschungsinstitut zur Zukunft der Arbeit (IZA). Bonn.
- Mikula, G. (1986). The experience of injustice: Toward a better understanding of its phenomenology. In H. W. Bierhoff, R. L. Cohen & J. Greenberg (Eds.), *Justice in social relations* (pp. 103-123). New York: Plenum Press.
- Mikula, G. (1993). On the Experience of Injustice. *European Review of Social Psychology*, 4, 223-244.
- Mikula, G., Scherer, K. R. & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin*, 24(7), 769-783.
- Miller, D. T. (2001). Disrespect and the experience of injustice. *Annual Review of Psychology*, 52, 527-553.
- Miller, D. T. & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, 74(1), 53-62.
- Miller, D. T. & Vidmar, N. (1981). The Social Psychology of Punishment Reactions. In M. J. Lerner & S. C. Lerner (Eds.), *The Justice Motive in Social Behavior*. New York: The Plenum Press.
- Mischel, W. (1976). *Introduction to Personality* (2. ed.). New York: Holt, Rinehart & Winston.
- Mischel, W. (1977). The Interaction of Person and Situation. In D. Magnusson & S. Endler (Eds.), *Personality at the Crossroads* (pp. 333-352). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Montada, L. (1998). Justice: Just a Rational Choice? *Social Justice Research*, 11(2), 81-101.

- Nikiforakis, N. S. (2004). *Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves?* Working paper. University of London, Royal Holloway, Department of Economics.
- Nowak, M. A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298.
- O'Gorman, R., Wilson, D. S. & Miller, R. R. (2005). Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior*, 26(5), 375-387.
- Oosterbeek, H., Sloof, R. & Van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2), 171-188.
- Ostrom, E. (2000). Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives*, 14(3), 137-158.
- Ottone, S. (2005). *Transfers and altruistic punishment in Solomon's Game experiments*. Working Paper. Nr.: 57. Università del Piemonte Orientale.
- Panchanathan, K. & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432(7016), 499-502.
- Pillutla, M. M., Malhotra, D. & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39(5), 448-455.
- Pillutla, M. M. & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208-224.
- Prasnikar, V. & Roth, A. E. (1992). Considerations of Fairness and Strategy - Experimental-Data from Sequential Games. *Quarterly Journal of Economics*, 107(3), 865-888.
- Preacher, K. J. & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods Instruments & Computers*, 36(4), 717-731.
- Rabin, M. (1993). Incorporating Fairness into Game-Theory and Economics. *American Economic Review*, 83(5), 1281-1302.

- Roth, A. E. (1995). Bargaining Experiments. In J. H. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 255-348). Princeton: Princeton University Press.
- Roth, A. E., Prasnikar, V., Okunofujiwara, M. & Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo - an Experimental-Study. *American Economic Review*, 81(5), 1068-1095.
- Roth, A. E. & Schoumaker, F. (1983). Expectations and Reputations in Bargaining - an Experimental-Study. *American Economic Review*, 73(3), 362-372.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755-1758.
- Schelling, T. C. (1978). Altruism, Meanness, and Other Potentially Strategic Behaviors. *American Economic Review*, 68(2), 229-230.
- Schlösser, T. (2006). *Against all odds? The role of emotions in decisions to trust or not to trust*. Unveröffentlichte Diplomarbeit. Universität zu Köln.
- Schmitt, M. (1996). Individual differences in sensitivity to befallen injustice (SBI). *Personality and Individual Differences*, 21(1), 3-20.
- Schmitt, M., Gollwitzer, M., Maes, J. & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment*, 21(3), 202-211.
- Schmitt, M., Maes, J. & Schmal, A. (1997). *Gerechtigkeit als inner-deutsches Problem: Analyse der Meßeigenschaften von Meßinstrumenten für Einstellungen zu Verteilungsprinzipien, Ungerechtigkeitssensibilität und Glaube an eine gerechte Welt*. Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral". Nr.: 105. Universität Trier, Fachbereich I - Psychologie.
- Schmitt, M. & Mohiyeddini, C. (1996). Sensitivity to Befallen Injustice and Reactions to a Real-Life Disadvantage. *Social Justice Research*, 9(3), 223-238.

- Schmitt, M., Neumann, R. & Montada, L. (1995). Dispositional Sensitivity to Befallen Injustice. *Social Justice Research*, 8(4), 385-407.
- Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung* (pp. 136-168). Tübingen: J.C.B. Mohr.
- Shinada, M., Yamagishi, T. & Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior*, 25(6), 379-393.
- Sigmund, K. & Hauert, C. (2002). Altruism. *Current Biology*, 12(8), R270-R272.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466-469.
- Skarlicki, D. P. & Kulik, C. T. (2005). Third-party reactions to employee (mis)treatment: A justice perspective. In *Research in Organizational Behavior: an Annual Series of Analytical Essays and Critical Reviews*, Vol 26 (Vol. 26, pp. 183-229). New York: Jai-Elsevier Science Inc.
- Smith, A. (1759/1982). *The Theory of Moral Sentiments*. Indianapolis: Liberty Fund.
- Sober, E. & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Tajfel, H., Billig, M. G., Bundy, R. P. & Flament, C. (1971). Social Categorization and Intergroup Behavior. *European Journal of Social Psychology*, 1(2), 149-177.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410-433.
- Trivers, R. L. (1971). Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46(1), 35-.

- Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E. & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes*, 89(1), 839-865.
- van't Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research*, 169(4), 564-568.
- Vidmar, N. (2001). Retribution and Revenge. In J. Sanders & V. L. Hamilton (Eds.), *Handbook of Justice Research in Law* (pp. 31-63). New York: Kluwer Academic / Plenum Publishers.
- von Neumann, J. & Morgenstern, O. (1944). *Theories of games and economic behavior*. Princeton: Princeton University Press.
- Walster, E., Walster, G. W. & Berscheid, E. (1978). *Equity: Theory and Research*. Boston: Allyn and Bacon.
- Westermarck, E. (1932). *Ethical Relativity*. New York: Harcourt, Brace.
- Wiswede, G. (1998). *Soziologie: Grundlagen und Perspektiven für den wirtschafts- und sozialwissenschaftlichen Bereich* (3. Aufl.). Landsberg am Lech: MI Verlag Moderne Industrie.
- Wiswede, G. (2001). Solidarität: Versuch einer Synopse. In H. W. Bierhoff & D. Fetchenhauer (Hrsg.), *Solidarität. Konflikt, Umwelt und Dritte Welt* (S. 323-350). Opladen: Leske + Budrich.
- Wiswede, G. (2007). *Einführung in die Wirtschaftspsychologie* (4. überarb. und erw. Aufl.). München: Ernst Reinhardt Verlag.
- Xiao, E. & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7398-7401.
- Yamagishi, T. (1986). The Provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology*, 51(1), 110-116.
- Zwick, R. & Chen, X. P. (1999). What price fairness? A bargaining study. *Management Science*, 45(6), 804-823.

## 8. Anhang

*Instruktionen und Items der Ungerechtigkeitssensibilität aus der Opfer-Perspektive (US<sub>Opfer</sub>) (Schmitt et al., 1997):*

Menschen reagieren in unfairen Situationen sehr unterschiedlich. Wie ist es bei Ihnen? Zunächst geht es um Situationen, die zum Vorteil anderer und zu **Ihrem** Nachteil ausgehen.

Bitte kreuzen Sie jeweils an, in welchem Ausmaß die Aussage auf Sie zutrifft. Die Antwortkategorien gehen von 0 („trifft überhaupt nicht zu“) bis 5 („trifft voll und ganz zu“).

1. Es macht mir zu schaffen, wenn andere etwas bekommen, was eigentlich mir zusteht.
2. Es ärgert mich, wenn andere eine Anerkennung bekommen, die ich verdient habe.
3. Ich kann es schlecht ertragen, wenn andere einseitig von mir profitieren.
4. Wenn ich Nachlässigkeiten anderer ausbügeln muss, kann ich das lange Zeit nicht vergessen.
5. Es bedrückt mich, wenn ich weniger Möglichkeiten bekomme als andere, meine Fähigkeiten zu entfalten.
6. Es ärgert mich, wenn es anderen unverdient besser geht als mir.
7. Es macht mir zu schaffen, wenn ich mich für Dinge abrackern muss, die anderen in den Schoß fallen.
8. Wenn andere ohne Grund freundlicher behandelt werden als ich, geht mir das lange durch den Kopf.
9. Es belastet mich, wenn ich für Dinge kritisiert werde, über die man bei anderen hinwegsieht.
10. Es ärgert mich, wenn ich schlechter behandelt werde als andere.



*Instruktionen und Items der Ungerechtigkeitssensibilität aus der Beobachter-Perspektive ( $US_{\text{Beobachter}}$ ) (Schmitt et al., 1997):*

Nun geht es um Situationen, in denen Sie mitbekommen oder erfahren, dass jemand **anderes** unfair behandelt, benachteiligt oder ausgenutzt wird.

Bitte kreuzen Sie jeweils an, in welchem Ausmaß die Aussage auf Sie zutrifft. Die Antwortkategorien gehen von 0 („trifft überhaupt nicht zu“) bis 5 („trifft voll und ganz zu“).

1. Es macht mir zu schaffen, wenn jemand nicht das bekommt, was ihm eigentlich zusteht.
2. Ich bin empört, wenn jemand eine Anerkennung nicht bekommt, die er verdient hat.
3. Ich kann es schlecht ertragen, wenn jemand einseitig von anderen profitiert.
4. Wenn jemand die Nachlässigkeiten anderer ausbügeln muss, kann ich das lange Zeit nicht vergessen.
5. Es bedrückt mich, wenn jemand weniger Möglichkeiten bekommt als andere, seine Fähigkeiten zu entfalten.
6. Ich bin empört, wenn es jemandem unverdient schlechter geht als anderen.
7. Es macht mir zu schaffen, wenn sich jemand für Dinge abrackern muss, die anderen in den Schoß fallen.
8. Wenn jemand ohne Grund freundlicher behandelt wird als andere, geht mir das lange durch den Kopf.
9. Es belastet mich, wenn jemand für Dinge kritisiert wird, über die man bei anderen hinwegsieht.
10. Ich bin empört, wenn jemand schlechter behandelt wird als andere.

*Instruktionen und Items der Ungerechtigkeitssensibilität aus der Nutznießer-/Täter-Perspektive (US<sub>Nutznießer</sub>) (Schmitt et al., 1997):*

Zuletzt geht es um Situationen, die zu **Ihren Gunsten** und zum Nachteil anderer ausgehen.

Bitte kreuzen Sie jeweils an, in welchem Ausmaß die Aussage auf Sie zutrifft. Die Antwortkategorien gehen von 0 („trifft überhaupt nicht zu“) bis 5 („trifft voll und ganz zu“).

1. Es macht mir zu schaffen, wenn ich etwas bekomme, was eigentlich anderen zusteht.
2. Ich habe ein schlechtes Gewissen, wenn ich eine Anerkennung bekomme, die andere verdient haben.
3. Ich kann es schlecht ertragen, wenn ich einseitig von anderen profitiere.
4. Wenn andere meine Nachlässigkeiten ausbügeln müssen, kann ich das lange Zeit nicht vergessen.
5. Es bedrückt mich, wenn ich mehr Möglichkeiten bekomme als andere, meine Fähigkeiten zu entfalten.
6. Ich habe Schuldgefühle, wenn es mir unverdient besser geht als anderen.
7. Es macht mir zu schaffen, wenn mir Dinge in den Schoß fallen, für die andere sich abrackern müssen.
8. Wenn ich ohne Grund freundlicher behandelt werde als andere, geht mir das lange durch den Kopf.
9. Es belastet mich, wenn man bei mir über Dinge hinwegsieht, für die andere kritisiert werden.
10. Ich habe Schuldgefühle, wenn ich besser behandelt werde als andere.

# Lebenslauf

## Franz N. Gresser

**Dipl. Volkswirt Franz N. Gresser**  
**Sülzgürtel 6**  
**50937 Köln**  
**Tel: 0221 – 4758468**  
**Mail: fng@gmx.de**

**Geburtsdatum: 22.07.1975 (Binningen, CH)**  
**Familienstand: ledig**

### **Ausbildung:**

Promotion (voraussichtlich Juli 2007)  
2003: Diplom Volkswirt (Note: 2.4), Universität zu Köln  
Vertiefung: Wirtschafts- und Sozialpsychologie  
Diplomarbeit: "Experimentelle Untersuchung des Einflusses von Persönlichkeitsvariablen auf das Risikoverhalten in Anlagesituationen" (Note: 1.0)  
1994: Abitur am Kreisgymnasium Kirchzarten (Note: 1.8; Englisch / Geschichte)  
1985-1994: Kreisgymnasium Kirchzarten  
1981-1985: Reinhold-Schneider-Grundschule, Freiburg

### **Berufliche Erfahrung:**

seit 06/2003: Wissenschaftlicher Mitarbeiter am Institut für Wirtschafts- und Sozialpsychologie der Universität zu Köln

### **Lehre:**

Wirtschaftspsychologische Seminare:  
Ökonomische Psychologie (WS 03/04 & WS 04/05)  
Psychologie der Entscheidung (WS 05/06 & WS 06/07)

### **Arbeit als studentische Hilfskraft:**

08/2000 – 05/2003: stud. Hilfskraft am Seminar für Wirtschafts- und Sozialpsychologie:  
Aufgabenfelder: VIRTUS (= Virtuelle Universitätssysteme), Institutshomepage, Layout verschiedener Veröffentlichungen, Mitarbeit am Sozialpsychologie-Lexikon von Prof. Wiswede

### **Praktika**

02/1997 - 03/1997 und 08.1997 – 09.1997  
BVU, Beratergruppe Verkehr und Umwelt GmbH in Freiburg

### **Zivildienst:**

08/1994-10/1995: Reha- und Kurklinik St. Urban in Freiburg

### **Ausgewählte Veröffentlichungen und Vorträge:**

- Gresser, F.N. & Fetchenhauer, D. (2006). Ungerechtigkeits-sensibilität und altruistische Bestrafung. Vortrag auf dem 45. Kongress der Deutschen Gesellschaft für Psychologie e.V. in Nürnberg
- Fetchenhauer, D., Gresser, F.N. & Haferkamp, A. (2006). Ein Volk von Betrügern? Schwarzarbeit, Steuerhinterziehung und Versicherungsmissbrauch aus psychologischer Sicht. In D. Enste & F. Schneider (Hrsg.). Jahrbuch Schattenwirtschaft 2006/2007 – Zum Spannungsfeld von Politik und Ökonomie. Berlin: LIT Verlag.
- Fetchenhauer, D., Gresser, F.N. & Haferkamp, A. (in press). Zwischen Zynismus und Gutgläubigkeit –Vertrauen wir anderen zu viel oder zu wenig? In F. Rexroth et al. (eds.). Kooperation und Betrug in Beziehungssystemen.
- Gresser, F.N. & Fetchenhauer, D. (2005). Moralisches Urteilen und altruistisches Bestrafen. Vortrag auf der 10. Tagung der Fachgruppe Sozialpsychologie, Jena. Gresser, F.N. & Fetchenhauer, D. (2005). The Psychology of Altruistic Punishment. Vortrag auf dem 30. IAREP Congress, Prag.
- Gresser, F.N., Krause, J.H. & Kiell, G. (2003). Glücksritter oder Rechenschieber? Welchen Einfluss haben Persönlichkeitsfaktoren und Anlagesituationen auf Finanzentscheidungen? Wirtschaftspsychologie, IV, 48-63.
- Gresser, F.N. (2003). Glücksritter oder Rechenschieber? Welchen Einfluss haben Persönlichkeitsfaktoren und Anlagesituationen auf Finanzentscheidungen? Vortrag im Rahmen des 3. Kölner Workshop "Quantitative Finanzmarktforschung".
- Gresser, F.N., Krause, J.H. & Kiell, G. (2002). Dispositionale und situative Determinanten des Risikoverhaltens bei Anlageentscheidungen. Vortrag in der Arbeitsgruppe Finanzpsychologie des 43. Kongress der Deutschen Gesellschaft für Psychologie, Berlin.

### **Sprachkenntnisse:**

Deutsch – Muttersprache  
Englisch – verhandlungssicher  
Französisch – gute Kenntnisse  
Spanisch – Grundkenntnisse

### **Computer- / EDV-Kenntnisse:**

SPSS, Unipark (Globalpark-Software für Online-Untersuchungen), Word, Excel, Powerpoint, HTML (Grundkenntnisse), CSS (Grundkenntnisse)